



Deskundigen over het testen van etnische minderheden

6. Arithmetic

Discontinue after 3 consecutive failures.

For ages 7–16, reverse sequence of preceding items after failure on either of first two items.

Bedeutung der IQ-Werte

Um die Erstellung von Gutachten, kann es sich erweisen, die numerischen IQ-Werte auch verbal zu interpretieren. Es gibt die diagnostischen Kategorien für verschiedene IQ-Bereiche. Diese wurden möglichst neutral gewählt und orientieren sich an denen der Originals.

	Time Limit	Comple. Time	Correct Response	Response	Score Circle one.	Prob
1. Count birds	30"		3		0 1	13. Jim
2. Count leaves	30"		2		0 1	14. Ne
3. Ice cream	30"		2		0 1	15. T-
4. Apple	30"		2		0 1	16. M
5. Ice cream	30"		2		0 1	17. De
6. Apple	30"		2		0 1	18. De

DESKUNDIGEN OVER HET TESTEN VAN ETNISCHE MINDERHEDEN

Redactie

mw.mr Najat Bochhah
drs. Willem Kort
drs. Hacene Seddik
prof.dr. Fons van de Vijver (eindredactie)

Met dank aan:
de geïnterviewde deskundigen
mw.drs. Anneke Bakker (bureauredactie)
mw.drs. Erna Kager (interviews)

Colofon

Uitgave van het

Landelijk Bureau ter bestrijding van Rassendiscriminatie
Schaatsbaan 51, 3013 AR Rotterdam,
Telefoon 010-2010201

en het

Nederlands Instituut van Psychologen
sectie Interculturalisatie
Postbus 9921, 1006 AP Amsterdam
Telefoon 020-4106220

Zet- en drukwerk:
Argus, drukken & ontwerpen
Rotterdam

Amsterdam, 2001

INHOUDSOPGAVE

Inleiding:

Deskundigen aan het woord over het gebruik van psychologische tests bij etnische minderheden	3
---	----------

Hoofdstuk

I	De conclusies, richtlijnen en aanbevelingen van de commissie-Hofstee	5
II	De Richtlijnen & Aanbevelingen: bekendheid, gebruik en tekortkomingen	7
III	Onderzoek	9
IV	Bestaande en nieuwe tests, hoe worden tests gekozen?	11
V	Richtlijnen	13
VI	Interpretatie van testresultaten	15
VII	Conclusie	17

Bijlage

1	Samenvatting interviews	19
2	Personalia van de geïnterviewde deskundigen	43
3	Vragenlijst 'Deskundigen over het testen van etnische minderheden'	45
4	Conclusies en aanbevelingen uit het rapport 'Toepasbaarheid van psychologische tests bij allochtonen' van de commissie-Hofstee	47
5	De tests van het rapport 'Toepasbaarheid van psychologische tests bij allochtonen' van de commissie-Hofstee	49
6	Overzicht van de in de interviews genoemde tests	51
7	Literatuur	53
8	Suggesties, per werkveld en algemeen	55

DESKUNDIGEN OVER HET TESTEN VAN ETNISCHE MINDERHEDEN

Inleiding:

Deskundigen aan het woord over het gebruik van psychologische tests bij etnische minderheden

Door het Landelijk Bureau ter bestrijding van Rassendiscriminatie (LBR) en het Nederlands Instituut van Psychologen (NIP) werd eind tachtiger jaren aan een commissie onder leiding van prof. Hofstee gevraagd om na te gaan of veel gebruikte psychologische tests eigenlijk wel geschikt waren voor toepassing bij allochtonen.

De commissie kwam in haar eindrapport met forse kritiek¹. Bestaande tests mochten dan wel geen racistische testinhoud hebben, goed bruikbaar voor allochtonen waren ze allerminst. Het complexe taalgebruik van veel tests en de veelvuldige verwijzingen naar elementen van de Nederlandse cultuur vormden forse bedreigingen voor een adequaat testgebruik bij allochtonen.

In de afgelopen tien jaar is de multiculturele samenstelling van de Nederlandse bevolking verder toegenomen. Zo is bijvoorbeeld de vraag naar psychologische expertise ten behoeve van vluchtelingen duidelijk gegroeid. Hoewel er geen exacte cijfers over zijn, valt aan te nemen dat het aantal geteste allochtonen in onderwijs, bedrijf en gezondheidszorg jaarlijks nog steeds toeneemt. De maatschappelijke relevantie van de beschikbaarheid van goede tests is dus niet minder geworden.

Het rapport-Hofstee kreeg bij het verschijnen nogal wat belangstelling, niet alleen bij vakgenoten. Het rapport heeft het testen van allochtonen uit het wetenschappelijke discours gehaald en het onderwerp bij een breder publiek bekend gemaakt. De belangstelling voor het onderwerp leek daarna weg te ebben. Maar is daarmee ook het probleem opgelost?

De vraag rijst dan hoe het tien jaar na het verschijnen van het rapport met de diagnostiek bij allochtonen gesteld is. Om een antwoord op deze vraag te krijgen hebben de sectie Interculturalisatie van het NIP en het LBR opnieuw het initiatief genomen een onderzoek te doen en wel onder experts op het terrein van multiculturele diagnostiek. Deze experts zijn zowel praktiserende diagnostici als (universitaire) onderzoekers. Aan hen zijn vragen voorgelegd als: wat is de invloed van het werk van de commissie-Hofstee? Zijn testcon-

structeurs, -afnemers en getesten op de hoogte van mogelijke problemen van multiculturele diagnostiek? Zijn nu voldoende tests beschikbaar? Zijn er praktische richtlijnen voor testgebruik?

Het hier gerapporteerde onderzoek heeft tot doel antwoord te krijgen op deze vragen. Het gebruik van psychologische tests is een activiteit op basis waarvan vaak vérstrekkende besluiten over personen genomen worden dan wel adviezen geformuleerd worden om deze besluiten te gaan nemen. Het is een wezenlijk onderdeel van het wetenschappelijk bedrijf om eigen activiteiten aan een breed publiek te presenteren en in het openbaar aan te geven hoe experts zelf tegen hun eigen activiteiten aankijken om aan de hand daarvan een zo goed mogelijk beeld van de bestaande situatie te verkrijgen.

Dit rapport is de weerslag van een inventarisatie zoals die in de zomer van 2001 uitgevoerd is. De gegevens die ten grondslag liggen aan dit verslag zijn verzameld in gesprekken met 15 deskundigen die direct betrokken zijn bij het testen en diagnosticeren van allochtonen. Zij zijn afkomstig uit drie werkvelden: Werving & Selectie, Onderwijs en Geestelijke Gezondheidszorg. Uit alle werkvelden zijn zowel praktiserende diagnostici als universitaire onderzoekers ondervraagd². Er is gewerkt met een standaardvragenlijst met open vragen³. De geïnterviewden waren vooraf op de hoogte van het onderwerp van het interview maar niet van de specifieke vragen; ze hebben derhalve “fris van de lever” geantwoord. Hun mening werd gevraagd over de conclusies, richtlijnen en aanbevelingen van de commissie-Hofstee, en de ontwikkelingen sindsdien op het vlak van onderzoek, revisies van bestaande tests en ontwikkeling van nieuwe, ontwikkeling van richtlijnen, interpretatie van gegevens en het gebruik van specifieke deskundigen daarbij.

Daarnaast is aan het eind van elk gesprek een open vraag gesteld. Men kon daar nog belangrijk geachte zaken inbrengen in toevoeging op de al gestelde vragen. Deze informatie is niet apart opgenomen maar vervlochten in de resultaten gepresenteerd in de verschillende hoofdstukken. De hoofdstukken in het verslag houden de volgorde van het interview aan. De door de auteurs geautoriseerde tekst van de volledige interviews (bijlage 1) is in de bijlagen opgenomen. Ook zijn alle tests (bijlage 6), onderzoeken en andere literatuur (bijlage 7) waar in de tekst van het rapport en in de interviews naar wordt verwezen opgenomen, evenals de diverse suggesties die de deskundigen hebben gedaan (bijlage 8).

¹ Zie bijlage 4 met de conclusies en aanbevelingen van de commissie-Hofstee, en bijlage 5 voor een overzicht van de door hen onderzochte tests.

² In bijlage 2 zijn de gegevens van de deskundigen opgenomen.

³ Zie bijlage 3.

I DE CONCLUSIES, RICHTLIJNEN EN AANBEVELINGEN VAN DE COMMISSIE-HOFSTEE

De gegevens in dit hoofdstuk zijn gebaseerd op de antwoorden op vraag 1⁴: Vindt u dat de conclusies van de commissie-Hofstee nog steeds gelden? Waar baseert u uw oordeel op?

De conclusies van de commissie-Hofstee waren – kort samengevat:

1. Er is geen racisme in testinhoud aangetroffen, maar ethnocentrisme ten gunste van de dominante cultuur is een bijna universeel verschijnsel in tests;
2. Vergelijkend onderzoek naar scores van allochtonen en autochtonen ontbreekt en dit beperkt de toepassing van tests sterk;
3. Onderzoek naar itembias en testbias ontbreekt nagenoeg in Nederland.

Deze conclusies worden in zijn algemeenheid door de geïnterviewden nog steeds onderschreven, met name door de deskundigen, werkzaam op het terrein van de Geestelijke Gezondheidszorg.

Ten aanzien van de eerste conclusie (de *ethnocentrische inhoud* van tests), wordt algemeen geconcludeerd dat er zo goed als niets veranderd is, in elk geval niet aan bestaande tests. De enige positieve uitzondering die genoemd wordt is de ontwikkeling van de nieuwe Multiculturele Capaciteiten Test – Middelbaar niveau (MCT-M) door Bleichrodt en Van den Berg, deze test is vrij van ethnocentrische inhoud. Met name de situatie op het gebied van persoonlijkheidstests acht men slecht: de beschikbare/bestaande tests zijn sterk verouderd en niet gereviseerd. Als voorbeeld wordt hierbij genoemd de zeer veel toegepaste Nederlandse Persoonlijkheids Vragenlijst (NPV), welke uit 1974 stamt en waar sterk cultuurgebonden items als “ik vind dat arme landen zichzelf moeten redden” en “volgens mij komen de beste bruggenbouwers uit ons land” nog steeds in voorkomen. Wel wordt er naar het oordeel van de in het wetenschappelijk onderwijs werkzame deskundigen bij de ontwikkeling van nieuwe tests beter gelet op de ethnocentrische valkuilen. Hiermee wordt met name

gedoeld op het taalgebruik bij de formulering van testitems en de instructie.

Taalgebruik is een belangrijke factor in ethnocentrische inhoud, een deskundige formuleerde het probleem als volgt: “... dubbele ontkenningen in een zin of antwoord, staande uitdrukkingen, gezegdes, beeldspraak, niet eenduidige formuleringen, verschillen in betekenissen van woorden. Het duurt heel lang voordat men een vreemde taal in nuances beheerst ...”. In alle door de commissie Hofstee onderzochte tests werd dergelijke ethnocentrische inhoud gevonden, en omdat geen van die tests gereviseerd is, is hierin niets veranderd.

Wat betreft de tweede conclusie (de *vergelijking van scores* van allochtonen en autochtonen) is met name op het terrein van Werving & Selectie wel het één en ander gebeurd. Zo is er onderzoek gedaan naar differentiële voorspelling en naar het onderscheidend vermogen van capaciteitentests. Uit het onderzoek naar differentiële voorspelling, waarin nagegaan is of capaciteitentests het feitelijk functioneren van allochtonen over- dan wel ondervoorspellen, is gebleken dat deze tests het feitelijk functioneren ook bij allochtonen goed voorspellen. De tests zijn dus wel eerlijk (hoewel niet altijd duidelijk is of metingen van het feitelijk functioneren in de vorm van beoordelingen door supervisors zelf wel zo’n goede maat zijn). In het onderzoek naar het onderscheidend vermogen, staat de vraag centraal of het onderscheid tussen hoog- en laagscorenden terecht is als gelet wordt op het praktisch functioneren. Daarbij is naar voren gekomen dat het onderscheidend vermogen van tests over het algemeen goed is en dat kandidaten met lagere scores in de praktijk ook daadwerkelijk minder goed functioneren dan kandidaten met hogere scores. Vaak kan het verschil verklaard worden door bijvoorbeeld discriminatie op de werkvloer.

De ontwikkelingen in de werkvelden Geestelijke Gezondheidszorg en Onderwijs blijven hierbij duidelijk achter. Er heeft op deze zeer belangrijke en omvangrijke werkerreinen tot dusver vrijwel geen enkel onderzoek plaatsgevonden.

⁴ Zie bijlage 3 voor de complete vragen.

Naast etnische achtergrond is door een aantal geïnterviewden ook gewezen op het belang van de sociaal-economische klasse voor het interpreteren van test-scores: mensen uit lagere sociaal-economische klassen scoren gemiddeld lager op tests en juist allochtonen zijn meer vertegenwoordigd in deze lagere klassen. De betreffende deskundigen zijn hier evenwel niet pessimistisch over: verwacht mag worden dat dit klas-senverschil bij de tweede en volgende generaties zal afnemen naarmate zij meer geïntegreerd zijn in de Nederlandse samenleving wat betreft taal, opleiding en werk.

Overigens is door bijna alle respondenten uitdrukke-lijk gewezen op onderzoekstechnische problemen bij vergelijking van testscores, met name waar het gaat om representativiteit en validiteit: de te onderzoeken groepen allochtonen zijn vaak te klein en heel ver-schillend van samenstelling, waardoor uitkomsten niet representatief (ze gelden niet voor de hele populatie) en niet valide (meet je wel wat je wilt meten) zijn.

Ook valt niet zomaar aan te nemen dat begrippen die van toepassing zijn op de autochtone populatie, met dezelfde onderzoeksinstrumenten te meten zijn bij een allochtone populatie (een probleem in de con-structvaliditeit van een test). Deze problemen zijn las-tig te ondervangen.

De derde conclusie heeft betreft het vrijwel geheel ont-breken van onderzoek naar item- en testbias in Nederland. Het gaat hier om het belangrijke probleem van mogelijke partijdigheid of vertekening. Hieronder wordt verstaan het ten onrechte benadelen van alloch-tone groepen door een test, testvragen of -opgaven.

De commissie-Hofstee wijst in dit verband op het vol-gende:

1. Er bestaat een aanwijzing voor partijdigheid als er sprake is van verschillen in gemiddelde testpres-taties, waarbij de testprestaties van vergelijkbare allochtone en autochtone groepen van elkaar ver-schillen ten nadele van allochtone groepen. De bewijslast dat de test fair is en niet-partijdig is zal bij de testgebruiker berusten.
2. Het is niet voldoende om aan te tonen dat een test niet ondervoorspelt. Er kan niet worden vol-gehouden dat een test fair is wanneer hij welis-waar niet ondervoorspelt, maar er sprake is van lagere arbeidsprestaties door bijvoorbeeld discrim-inatie of lagere schoolprestaties door onvol-doende goed onderwijs.
3. Er is geen sprake van partijdigheid als er geen ondervoorspelling is en kan worden aangetoond dat het testcriterium geen bias vertoont.

Door veel deskundigen is er op gewezen dat de laat-ste jaren wel veel onderzoek naar bias is verricht, waarbij de technieken ook zeer aanzienlijk zijn verbe-terd. Daarbij sluiten de resultaten van Nederlands onderzoek goed aan bij elders – met name in de Verenigde Staten – verricht onderzoek, met als belang-rijke conclusie dat mogelijke partijdigheid op het niveau van testvragen of –opdrachten bij cognitieve taken (intelligentie) weinig invloed heeft op de ver-schillen in scores tussen allochtonen en autochtonen. Bij persoonlijkheidstests is mogelijk meer vertekening. Overigens wordt door de in de praktijk werkzame des-kundigen dit punt vooral een academische kwestie genoemd, dat niet erg sterk leeft in de professionele psychologiebeoefening.

CONCLUSIE

Conclusie

Samenvattend kan worden gesteld dat de conclusies van de commissie Hofstee in zijn algemeenheid nog steeds gel-dig zijn en dat er sinds het verschijnen van het rapport nauwelijks of geen wezenlijke verbeteringen tot stand zijn gekomen. Wel is er enig onderzoek verricht naar differentiële voorspellingsmogelijkheden, met name op het terrein van Werving & Selectie (capaciteitentests) en ook is in onderzoek redelijk veel aandacht geschonken aan het pro-bleem van de partijdigheid. Het onderzoek is echter over het algemeen nogal academisch van karakter en op belang-rijke toepassingsgebieden als het onderwijs en de (geestelijke) gezondheidszorg is vrijwel geen voortgang gemaakt.

II DE RICHTLIJNEN & AANBEVELINGEN: BEKENDHEID, GEBRUIK EN TEKORTKOMINGEN

De gegevens in dit hoofdstuk zijn gebaseerd op de antwoorden de vraag 2⁵: Welke richtlijnen en aanbevelingen van de commissie-Hofstee kent u, en welke zijn – naar uw mening – bekend bij testconstructeurs, testgebruikers en getesten? Spelen ze een rol bij uw testkeuze en die van anderen? Wat zijn tekortkomingen ervan en hoe kunnen deze worden opgeheven?

Bekendheid

Bij geïnterviewden zelf

Alle geïnterviewden waren in mindere of meerdere mate bekend met de richtlijnen en aanbevelingen, alhoewel niet in detail en zonder deze alle expliciet te kunnen aangeven. Uiteraard hangt e.e.a. samen met de mate waarin men bij het rapport betrokken is geweest: enkele geïnterviewden hebben zelf meegewerkt aan het tot stand komen van het rapport en zijn er derhalve zeer goed mee bekend. Daarnaast blijken ook de geïnterviewden die lid zijn van de Commissie Test Aangelegenheden (COTAN⁶) van het NIP goed op de hoogte te zijn van de richtlijnen, omdat de COTAN in haar eigen richtlijnen uitdrukkelijk rekening heeft gehouden met de bevindingen van de Commissie.

De geïnterviewden noemden als belangrijkste thema's uit de richtlijnen en aanbevelingen van de commissie-Hofstee:

- ▶ "Niet-racistische inhoud".
- ▶ "Wees voorzichtig met het testen van allochtonen i.v.m. taalvaardigheid en culturele misperceptie van items, ga niet klakkeloos uit van de validiteit van de test, en verifieer de interpretatie met aanvullende informatie".
- ▶ "De validiteit en de betrouwbaarheid van de tests is in het geding. Er is tot op heden weinig onderzoek naar gedaan. Van iedere willekeurige test kunnen we stellen dat er geen racistische intentie in de test zit. In ieder geval is dat de bedoeling. Het gaat veel te ver om te stellen dat dit zou zijn".

Men is zeer betrokken bij het onderwerp in algemene zin en tracht zich daarbij een zo goed mogelijk beeld te vormen van alle aspecten die daarmee samenhangen. Het rapport van de commissie wordt daarbij van groot belang geacht en voor het overige heeft men de bevindingen aangevuld met andere informatie die sindsdien in publicaties is verschenen.

Bij anderen

Men is unaniem van oordeel dat testconstructeurs redelijk op de hoogte zijn van de aanbevelingen en dat er ook rekening mee wordt gehouden bij het maken van tests. Voor wat betreft de bekendheid van de richtlijnen bij testgebruikers klinkt er duidelijk meer twijfel door. Men is er allesbehalve zeker van dat men in de praktijk voldoende op de hoogte is van de problematiek zoals die door de commissie naar voren is gebracht.

De vraag of de getesten de richtlijnen kennen, wordt door allen duidelijk ontkennend beantwoord. Voor zover bekend wordt bij voorkomende protesten tegen testprocedures of testuitslagen, bijvoorbeeld op het gebied van Werving & Selectie, nooit een expliciet beroep gedaan op de richtlijnen van de commissie. Na het verschijnen van het rapport stond het onderwerp even in de maatschappelijke belangstelling, maar deze is weggeëbd. Het onderwerp leeft nu niet meer.

Gebruik

Door geïnterviewden zelf

De deskundigen die lid zijn van de COTAN stellen dat de richtlijnen duidelijk een rol spelen in het beoordelen van nieuwe tests. Voor zover de geïnterviewden werkzaam zijn in het universitaire onderwijs geven ze aan dat zij in hun onderwijs uitdrukkelijk aandacht geven aan de richtlijnen en deze aan studenten overdragen.

Bij de overige (vooral de in de praktijk werkzame) deskundigen spelen de richtlijnen niet expliciet een rol, maar gaat het meer om algemene kennis waar rekening mee wordt gehouden in de praktische func-

⁵ Zie bijlage 3 voor de complete vragen.

⁶ De COTAN houdt zich bezig met de kwaliteitsbewaking van tests in het Nederlands Taalgebied, o.a. door het publiceren van Documentatie van tests en testresearch in Nederland (2000).

tie-uitoefening. Men richt zich hierbij op alle beschikbare informatie afkomstig uit de verschillende publicaties over het onderwerp.

In de dagelijkse praktijk

In navolging van de geconstateerde onbekendheid van de richtlijnen bij testgebruikers constateert men ook dat testgebruikers in de praktijk bepaald niet altijd even zorgvuldig optreden en niet of nauwelijks rekening houden met de aanbevelingen.

In dit opzicht is men weinig optimistisch over de wijze waarop in de praktijk tests worden toegepast. Zo merkt één van de universitaire onderzoekers op: "... van studenten die stage lopen, hoor ik dat er in de praktijk weinig terughoudendheid wordt betracht in interpretatie van testresultaten van allochtonen. De tests worden vaak gebruikt als heilige schrift." en: "In de handleiding staat een groep aangegeven maar in de praktijk wordt de test ook voor andere groepen gebruikt. Men is zich niet bewust van de risico's."

Tekortkomingen

Naar het oordeel van de deskundigen is er een aantal duidelijke tekortkomingen in het rapport die een goede doorwerking er van in de praktijk belemmeren. Als zodanig worden aangegeven:

- ▶ De uitkomsten in het rapport zijn te zeer gebaseerd op een inhoudsanalyse van vragenlijsten. Een empirische, kwantitatieve analyse van testgegevens ontbreekt.
- ▶ Richtlijnen over de keuze voor een test in een specifieke situatie worden gemist, de richtlijnen zijn zeer algemeen van karakter en missen een duidelijk toespitsing op specifieke situaties en omstandigheden.
- ▶ Het rapport heeft te weinig invloed gehad op de opinievorming. Toen het rapport verscheen is er veel aandacht aan besteed en was er belangstelling voor het onderwerp, maar uiteindelijk heeft het weinig effect gehad op de testpraktijk en zijn zeker nieuwe lichten testgebruikers slecht op de hoogte.
- ▶ De richtlijnen zijn vooral gericht op één werkveld: Werving & Selectie.

CONCLUSIE

Conclusie

De geïnterviewden zelf, zijn redelijk tot goed op de hoogte van problematiek en van de conclusies en aanbevelingen van de commissie-Hofstee. Men houdt duidelijk rekening met de richtlijnen of met de strekking daarvan bij de constructie en beoordeling van nieuwe tests en bij het onderwijs over diagnostiek. Het gaat hier dan ook om een groep met een specifieke deskundigheid op het terrein van testgebruik bij allochtonen. Van testconstructeurs schat men in dat deze dezelfde houding aannemen en voorzichtigheid betrachten.

Wat betreft de betekenis van de richtlijnen van de commissie voor de dagelijkse praktijk komt een somber beeld naar voren. Weliswaar heeft er een zekere bewustwording plaatsgevonden, maar toch kan niet gesproken worden van een duidelijke verandering in de wijze waarop tests worden toegepast. Voor een deel is dat toe te schrijven aan het vrij algemene karakter van de richtlijnen, waarbij weinig specifieke aanwijzingen zijn gegeven hoe in concrete en specifieke situaties met de richtlijnen het best kan worden omgegaan. Verder is van belang dat de richtlijnen nogal sterk lijken te zijn toegesneden op de Werving- & Selectiepraktijk en nauwelijks aansluiten bij de situatie zoals die binnen terreinen als Onderwijs en Geestelijke Gezondheidszorg bestaat.

Tenslotte lijken de richtlijnen nauwelijks enige maatschappelijke betekenis te hebben gehad: kandidaten of cliënten die getest worden lijken in het geheel niet op de hoogte te zijn en van enige maatschappelijke discussie over dit onderwerp kan geen aanduiding worden gevonden.

De gegevens in dit hoofdstuk zijn gebaseerd op de antwoorden op vraag 3⁷: Welk verder onderzoek is er sinds het rapport verricht? Spelen de uitkomsten van dit onderzoek een rol in de praktijk? Is dit onderzoek voldoende? Zo nee, waar is dat aan te wijten en hoe kan dit worden bevorderd?

Resultaten van onderzoek

Van alle onderzoeken die door de deskundigen worden aangehaald, worden de proefschriften van Te Nijenhuis (1997), Van Leest (1997) en Van den Berg (2001) het meest genoemd⁸. Deze zijn een gevolg geweest van het verschijnen van het rapport van de commissie-Hofstee.

Van deze onderzoeken heeft men als algemeen beeld dat cognitieve (intelligentie- en capaciteiten-) tests goed toepasbaar zijn en een reëel beeld van de cognitieve vermogens geven. Zo weet men van het onderzoek van Te Nijenhuis dat intelligentie- en capaciteitentests meestal over een goede betrouwbaarheid en validiteit beschikten, dat zij geen partijdige resultaten opleverden en goed toepasbaar zijn. Over de onderzoeken van Van den Berg en Van Leest weet men dat er betere instructies aanbevolen werden, en dat gewenning aan testsituaties, taal en het ontwikkelen van een strategie onder tijdsdruk belangrijke factoren zijn in de testuitslag.

Andere onderzoeksuitkomsten die men noemt betreffen de aanbeveling van het gebruik van andersoortige tests (b.v. videotests), het gebruik van meerdere informatiebronnen tegelijkertijd en de gebleken geringe invloed van testbias en itembias.

Bovengenoemde onderzoeken hadden vooral betrekking op de problematiek van meting van intelligentie en vaardigheden en zijn daarom ook het meest van belang voor het terrein van Werving & Selectie. Onderzoek naar mogelijkheden van toepassing en ontwikkeling van tests die gebruikt worden in de Geestelijke Gezondheidszorg – persoonlijkheidstests of gedragsproblematiek– ontbreekt geheel.

⁷ Zie bijlage 3 voor de complete vragen.

⁸ Zie voor deze en alle andere genoemde publicaties de literatuurlijst in bijlage 6.

Beperkingen

Naar het oordeel van vrijwel alle geïnterviewden is het uitgevoerde onderzoek te eenzijdig en te beperkt in omvang geweest. Bepleit wordt om niet uitsluitend aandacht te schenken aan de testtheoretische en psychometrische aspecten maar juist aandacht te schenken aan het gehele proces, waarbij de keuze, de wijze van gebruik, de interpretatie en het interactieproces met de geteste aan de orde komen. Juist op het terrein van persoonlijkheidsonderzoek is de behoefte aan meer inzicht in deze aspecten zeer duidelijk aanwezig. In dit opzicht is overigens een belangrijk punt van zorg het sterk toegenomen gebruik van tests in uiteenlopende maatschappelijke sectoren, waarbij vrijwel uitsluitend af wordt gegaan op testcores zonder dat men zich bewust lijkt te zijn van de achtergrond van de testuitslag en bruikbaarheid van de betreffende test. Vanuit een duidelijke bezorgdheid over deze situatie wordt algemeen gepleit voor de beschikbaarheid van goede deugdelijke tests, waarvan de toepassingsmogelijkheden en reikwijdte duidelijk zijn aangegeven. Tenslotte wordt benadrukt dat testgebruikers zich bij hun onderzoek niet alleen op tests richten maar op alle te gebruiken methodes en het proces.

De praktische betekenis van het onderzoek

In brede kring is men zich wel bewust geworden van de problematiek: er is sprake van een redelijke mate van voorzichtigheid in de toepassing van tests en interpretatie van resultaten. De onderzoekers onder de geïnterviewden stellen dat veel van het goede nieuws dat uit het onderzoek naar voren is gekomen, in het bijzonder wat betreft de geringe mate van partijdigheid, niet doordringt in de praktijk. De geïnterviewde deskundigen uit de praktijk beamen dit.

In dit verband wordt met name door de deskundigen uit het terrein van het Onderwijs expliciet aangegeven dat er geen brug geslagen is naar de praktijk. Het verrichte onderzoek heeft vrijwel uitsluitend in wetenschappelijke kring aandacht gekregen en heeft tot een sterk theoretische en academisch getinte discussie geleid. Het onderzoek sluit nauwelijks aan bij de problematiek zoals die in de dagelijkse praktijk naar voren komt. In dat opzicht heeft het onderzoek dan ook niet veel betekenis gehad en is er sprake van zeer weinig doorwerking. De conclusies van de commissie zijn in dat opzicht nog steeds actueel te noemen.

Daarnaast brengen vooral de professioneel werkzame deskundigen naar voren dat psychologisch onderzoek veel breder zou moeten worden opgevat dan alleen

maar het afnemen van tests. Zij constateren dat testgebruikers zich redden met wat er is, door een eigen persoonlijke testbenadering bij elkaar te sprokkelen. Men behelpt zich op deze wijze maar betreurt dat weinigen de moeite nemen om zich sterk te maken om veranderingen door te voeren en de bestaande, onbevredigende situatie te verbeteren.

In aansluiting hierop is een aantal onderwerpen aan-gegeven waar toekomstig onderzoek zich vooral op zou kunnen richten en waar vanuit de praktische beoefening van de diagnostiek in sterke mate behoefte aan bestaat. Genoemd zijn:

- ▶ Onderzoek naar mogelijkheden van testgebruik bij vluchtelingen. Het betreft hier een zeer specifieke groep met een duidelijk eigen problematiek: “Als je het bij vluchtelingen kunt onderzoeken, dan kun je het zeker bij andere allochtonen”.
- ▶ Onderzoek naar de toepasbaarheid van specifieke tests in verschillende situaties: “Op het ogenblik is de situatie zo, dat iedereen zelf het wiel moet gaan zitten uitvinden en dat is natuurlijk een verspilling van tijd en energie”.
- ▶ Longitudinaal onderzoek naar de mate van acculturatie.
- ▶ De wijze waarop doorverwijzing plaatsvindt op basis van testresultaten.
- ▶ Onderzoek naar de bruikbaarheid en toepassingsmogelijkheden van persoonlijkheidstests en ontwikkeling van nieuwe persoonlijkheidstests.

Wel wordt opgemerkt dat de verschillen binnen etnische groepen groot zijn, en dat dit algemene uitspraken en algemeen onderzoek bemoeilijkt.

Op de vraag waarom tot dusverre nauwelijks of geen onderzoek met betrekking tot de genoemde onder-

werpen is verricht, worden uiteenlopende antwoorden gegeven. Genoemd worden onder meer:

- ▶ *onderzoekstechnische* problemen (representativiteit, criteriumbias, de mate waarin men vertrouwd is met tests en testgebruik, sociaal-economische klasse als contaminerende factor);
- ▶ *gebrek aan tijd en geld*;
- ▶ *praktische drempels* met betrekking tot *geheimhouden van informatie* over normgroepen (die bij de grote Werving- & Selectiebureaus immers een deel van het bedrijfskapitaal vormen en daarom bedrijfsgeheim zijn);
- ▶ *financiering* van onderzoek vormt een groot probleem: Om gelden vrij te maken dienen de krachten te worden gebundeld en is het van belang dat universiteiten aan dergelijk onderzoek prioriteit geven. Met name in de Geestelijke Gezondheidszorg en het Onderwijs stuit men op deze problematiek. Op het terrein van Werving & Selectie is er een economisch belang mee gediend en gebeurt er meer.

Hoe te bevorderen?

Bepleit wordt het tot stand komen van een onafhankelijk samenwerkingsverband met diverse partijen (universiteit, praktijk, testverkopers, overheid), die onderzoek op dit vlak financiert én evalueert. Gezien de mogelijk politieke lading van dergelijk onderzoek wordt gesuggereerd daarin wel zeer zorgvuldig te werk te gaan: “Het NIP en de COTAN zouden hier een voortrekkersrol kunnen spelen. Het moet wel een neutrale partij zijn, anders trek je het te veel in een bepaalde hoek, krijg je een politieke lading”.

CONCLUSIE

Conclusie

Er is in de afgelopen tien jaar nogal wat onderzoek gedaan (vooral promotieonderzoek). In het boek van Bleichrodt en Van de Vijver (2001) is een overzicht te vinden. Het onderzoek is nuttig maar wat eenzijdig en er moet nog een lange weg afgelegd worden voordat multiculturele diagnostiek echt een volwassen vak geworden is.

De praktiserende diagnosticus vindt onvoldoende handvatten in bestaande onderzoekingen naar partijdigheid om de problemen van alledag in het testen van allochtonen mee te lijf te gaan. Verder is er vooral op het gebied van persoonlijkheid nog veel werk te doen. Voor bestaande tests die in het onderzoek niet goed bruikbaar bleken, zijn nog geen alternatieven ontwikkeld.

Het thema van een verdere professionalisering van multiculturele diagnostiek is een regelmatig terugkerend thema in de gesprekken. Omdat multiculturele diagnostiek zowel onderzoekers als praktische diagnostici bezig houdt, is het een goed thema om een dialoog tussen beide velden te initiëren. Het doen van door de praktijk geïnspireerd onderzoek, informatieverstrekking over bruikbaarheid van tests en het nemen van initiatieven tot het komen van een kwaliteitssysteem voor bruikbaarheid van psychologische tests in de multiculturele samenleving zijn slechts enkele thema's die in een gecombineerde groep van onderzoekers en practici aan bod zouden kunnen komen. Een respondentopperde de suggestie om het NIP en/of de COTAN hierin een initiërende en coördinerende rol te laten spelen.

De gegevens in dit hoofdstuk zijn gebaseerd op de antwoorden op vraag 4⁹: Zijn er sinds het rapport tests gereviseerd of nieuwe ontwikkeld? Zo nee, waarom niet? Zo ja, wordt er dan voldoende aangegeven in hoeverre er sprake is van ethnocentrisme? Houdt u zelf rekening met bestaand ethnocentrisme in een test? Hoe doen anderen dit op uw werkterrein?

Revisie van bestaande tests en ontwikkeling van nieuwe

In aansluiting op hetgeen hiervoor naar voren is gekomen wordt door de geïnterviewden breed geconstateerd dat er sinds de verschijning van het rapport van de commissie zeer weinig revisies van tests hebben plaatsgevonden en dat er eveneens weinig nieuwe tests zijn ontwikkeld.

Door bijna alle deskundigen wordt de Multiculturele Capaciteiten Test – Middelbaar niveau (MCT-M) genoemd als een nieuwe test op het gebied van intelligentie- en capaciteitentests. Deze is ontwikkeld door de aan de Vrije Universiteit Amsterdam gelieerde Stichting NOA.

Daarnaast worden veel verschillende tests genoemd die niet alle nieuw of gereviseerd zijn, maar bij voorzichtige toepassing wel bruikbaar zijn. De volgende tests werden door twee of meer mensen genoemd:

- ▶ Differentiële Aanleg Test (DAT),
- ▶ Leertest voor Etnische Minderheden (LEM), en
- ▶ Snijders-Oomen Niet-Verbale Intelligentietest (SON-R).

De geïnterviewden van de Werving- & Selectiebureaus melden verder dat zij eigen testmateriaal hebben ontwikkeld, zoals taaltests voorafgaand aan een assessment center of tests die gebruik maken van (aangepast) beeldmateriaal.

Met name op het terrein van de Geestelijke Gezondheidszorg zijn de ontwikkelingen zeer beperkt

geweest. Genoemd worden alleen de ontwikkeling van de NEO-PI en de NEO-FFI, welke tests hun waarde hebben bewezen in zeer uiteenlopende culturen.

Overigens wordt bij nieuwe tests naar het oordeel van vrijwel alle deskundigen niet of nauwelijks aangegeven of en in hoeverre er sprake zou kunnen zijn van ethnocentrisme. Slechts bij enkele tests, met name op het gebied van intelligentie, wordt daar aandacht aan geschonken.

Toepasbaarheid

Voor wat betreft de keuze van tests laat men zich in de praktijk primair leiden door de specifieke vraagstelling en de situatie van degene die getest wordt. In dat opzicht is de keuze van tests sterk door de vraagstelling en de concrete omstandigheden bepaald, waarbij de mogelijke geschiktheid van de test voor allochtone groepen slechts één van de overwegingen is.

Zo wordt in situaties waarbij de geteste nog maar kort in Nederland verblijft vooral gelet op het beroep dat de test op taalvaardigheid doet. In die situatie zijn tests met een sterk verbaal karakter niet te gebruiken, of zijn tests überhaupt niet te gebruiken.

Ook bij de opzet van een assessment center wordt zoveel mogelijk gelet op de persoonlijke omstandigheden en kan worden besloten om geen tests af te nemen of alleen een gestructureerd interview.

Verder tracht men in de praktijk zoveel mogelijk rekening te houden met mogelijk bestaande verschillen in de meting van te onderscheiden factoren. Zo wordt er veelal van uitgegaan dat de wijze waarop bijvoorbeeld verbale en performale intelligentie wordt gemeten waarschijnlijk in verschillende mate door ethnocentrisme zijn vertekend: “Daar hou je rekening mee. Als een leerling weinig Nederlands spreekt kijk je meer naar de performale en abstracte onderdelen”. Daarnaast tracht men ook uit andere bronnen gegevens te verzamelen en bij de interpretatie van de testresultaten na te gaan in hoeverre de verschillende gegevens met elkaar overeenstemmen.

⁹ Zie bijlage 3 voor de complete vragen.

Testconstructeurs streven, met het oog op een goede toepassing, ernaar om instructies en testaanwijzingen zo eenvoudig, simpel en duidelijk mogelijk te maken. Docenten op universiteiten wijzen ook uitdrukkelijk op de complicaties die het testen van allochtonen met zich meebrengt; studenten worden gewezen op specifieke tests die in dat verband bruikbaar zijn.

Ten aanzien van de vraag in hoeverre de geschetste handelwijze voorkomt en de vraag of ethnocentrisme een duidelijke invloed heeft op de keuze van tests, blijkt er vrijwel geen inzicht te bestaan in de feitelijke situatie. Het is vrijwel onbekend en er kan slechts de hoop worden uitgesproken dat men zich voldoende rekenschap geeft van het probleem en tot een verantwoorde keuze van tests komt

CONCLUSIE

De situatie wat betreft de ontwikkeling en revisie van tests wordt als weinig rooskleurig beschreven. Bij de ontwikkeling van tests wordt nauwelijks aandacht besteed aan de vraag of er sprake is van ethnocentrisme en er is sinds het verschijnen van het rapport van de commissie slechts één test verschenen die daar vrij van is. Verder is er hier en daar materiaal ontwikkeld voor eigen gebruik, met name gericht op het vaststellen van de taalvaardigheid.

Tegen deze achtergrond kunnen slechts enkele tests als redelijk tot goed bruikbaar worden geacht, mits er bij de toepassing de nodige omzichtigheid in acht wordt genomen. Hierbij betreft het meestal capaciteiten- of intelligentietests en enkele persoonlijkheidstests. Tests ten behoeve van specifieke toepassingen op het terrein van de Geestelijke Gezondheidszorg of binnen het Onderwijs die vrij zijn van ethnocentrisme zijn in het geheel niet beschikbaar.

In de praktijk laat men de keuze van de te gebruiken middelen vooral afhangen van de concrete situatie, waarbij een zo goed mogelijke afweging van de verschillende aspecten wordt gemaakt. De kwaliteit van de test is daarbij slechts één van de overwegingen. Voor het overige lijkt men zich vooral door de eigen kennis en ervaring te laten leiden.

Een belangrijk aandachtspunt is de schatting van de taalvaardigheid, en afhankelijk daarvan wordt vaak beslist of bepaalde verbale (subtests) worden gebruikt. Er wordt daarbij – al dan niet stilzwijgend – van uit gegaan dat de mate van vertekening op performante (sub)tests minder zou zijn dan bij verbale subtests. Daarnaast neemt men vaak de toevlucht tot mogelijke beschikbare andere informatie en tracht die als basis voor de oordeelsvorming te gebruiken.

De gegevens in dit hoofdstuk zijn gebaseerd op de antwoorden op vraag 5¹⁰: Zijn in de afgelopen jaren voldoende goede en bruikbare richtlijnen ontwikkeld? Welke zijn dit, en zijn deze algemeen toepasbaar? Houdt u er rekening mee in uw oordeelsvorming? Als ze niet bestaan, moeten deze dan worden ontwikkeld en aan welke eisen moeten deze dan voldoen?

Nieuwe richtlijnen?

Geïnterviewden uit de drie werkvelden constateren dat er sinds het verschijnen van het rapport in 1990 geen nieuwe richtlijnen zijn ontwikkeld.

Wel achten alle geïnterviewden het wenselijk dat er goede richtlijnen komen en allen hameren daarbij op hetzelfde punt: de algemene toepasbaarheid. Verder zal in de handleiding van elke test moeten worden aangegeven of er sprake is van partijdigheid. Dergelijke nieuwe richtlijnen zouden gekoppeld dienen te worden aan de richtlijnen van de commissie-Hofstee. De *Standards for Educational and Psychological Testing* van de American Educational Research Association, American Psychological Association en National Council on Measurement in Education (1999) worden genoemd als een goed voorbeeld.

Sommige deskundigen kunnen redelijk uit de voeten met hetgeen er is ontwikkeld aan richtlijnen; ondanks het feit dat deze meestal afkomstig zijn uit de Werking & Selectie en loopbaanbegeleiding, zijn ze wel algemeen toepasbaar.

Er zijn concrete ideeën over de vorm van dergelijke nieuwe richtlijnen. Hierbij wordt niet alleen aangegeven aan welke zorgvuldigheidseisen ten aanzien van de testkwaliteit en –toepassing moet worden voldaan, maar wordt het gehele proces (van testconstructie tot en met interpretatie van scores) zoveel mogelijk gespecificeerd aan de hand van kwaliteitseisen: “Er zou een brochure moeten komen over welk instrument wanneer te gebruiken is, met regelmatige updates, vanuit de praktijkervaringen. Zoals de huisartsen protocollen hebben omtrent klachten en wat daar verder aan te doen. Dan krijg je veel meer sturing in het gebruik van tests”.

Naast richtlijnen voor de adequate toepassing van bestaande en nieuwe tests wordt aandacht gevraagd voor mogelijkheden om op een andere wijze tot oordeelsvorming te komen. Hiermee wordt met name bedoeld op een meer procesmatige benadering, waarbij de testuitslag aanknopingspunten biedt voor een verdere oordeelsvorming, waarbij ook andere gegevens worden meegewogen. Het opstellen van richtlijnen voor juist testgebruik moet niet leiden tot het uitsluiten van andere beoordelingswijzen.

CONCLUSIE

Conclusie

Er is duidelijk behoefte aan goede en algemeen toepasbare richtlijnen, die concreet howvast kunnen bieden voor de praktijk. Dergelijke richtlijnen dienen aan te sluiten bij de richtlijnen zoals de commissie-Hofstee die heeft geformuleerd en zouden naar het voorbeeld van bestaande standaarden (bijvoorbeeld welke in de Verenigde Staten bestaan) kunnen worden opgezet.

In dat opzicht is ook aansluiting te zoeken bij de aanpak zoals die wordt toegepast bij de formulering van een Best Test Practice¹¹. Juist een dergelijke aanpak zou het mogelijk maken om het gehele proces van constructie tot keuze, toepassing en rapportage zo goed mogelijk te laten verlopen en daarbij aan te sluiten bij de situatie zoals die in de concrete praktijk bestaat en daar kwaliteitsnormen aan te koppelen.

Overigens wordt het van belang geacht dat ook andere gegevens dan alleen testresultaten in de oordeelsvorming worden betrokken.

¹⁰ Zie bijlage 3 voor de complete vragen.

¹¹ Zie voor meer informatie: <http://www.4tp.dyndns.org>

De gegevens in dit hoofdstuk zijn gebaseerd op de antwoorden op vragen 6 en 7¹²: Bent u terughoudend bij de testinterpretatie van allochtone kandidaten? Zo ja, hoe komt dit tot uiting, en is dit voldoende? Hoe doen anderen in uw werkveld dit, en is dit voldoende? Zijn er specifieke deskundigen nodig?

Terughoudendheid in interpretatie

Bij geïnterviewden zelf

Eén van de meest gehanteerde tactieken om de nodige voorzichtigheid in de interpretatie bij het testen van allochtonen in te bouwen, is overleg met collega's, al dan niet van allochtone afkomst. Juist een allochtone collega wordt in staat geacht om "van binnenuit de problematiek te begrijpen". Voor de testgebruiker zijn ook "contacten met andere culturen, verblijf in het buitenland, geschoold in een interculturele setting" hulpmiddelen.

Verder wordt benadrukt dat het belangrijk is om niet alleen af te gaan op de tests en de scores maar om zoveel mogelijk informatie te verzamelen (ook uit andere bronnen) om tot een goed afgewogen oordeel te kunnen komen. Deze manier van werken hanteert men overigens bij alle beoordelingen, niet alleen waar het allochtonen betreft. Alternatieve informatiebronnen die in dit verband genoemd worden zijn:

- ▶ simulatieoefeningen;
- ▶ voorbereiding en begeleiding kandidaat, brede samenstelling assessment, rollenspellen, interviews;
- ▶ inzicht in de beheersing van de eigen taal van de geteste;
- ▶ oordeel leerkracht;
- ▶ observatie;
- ▶ projectief narratief materiaal;
- ▶ informatie uit zoveel mogelijk bronnen: dossier, tenlastelegging, familie, sociale omgeving, persoonlijke historie, heteroanamnese.

Taal speelt ook een belangrijke rol: eenvoudig taalgebruik, eenduidigheid in begrippen, geen gezegdes, beeldspraak, geen dubbele ontkenningen. Taal speelt tevens een rol in de omgang met de te testen persoon. Het is belangrijk het vertrouwen van de persoon in kwestie te winnen, bijvoorbeeld door het kennen van enkele woorden uit diens eigen taal.

De geïnterviewden constateren algemeen dat er geen cultuurvrije tests te ontwerpen zijn, en dat men in een aantal gevallen naar eigen inzicht handelt. Bij de interpretatie zal daarom terughoudendheid altijd een belangrijk aandachtspunt blijven.

Bewustzijn van de eigen cultuur bij testgebruikers is een voorwaarde, maar niet afdoende om de problemen te verhelpen.

Bij anderen

Door de mensen uit de praktijk van de Werving & Selectie wordt melding gemaakt van het feit dat – met name bij de kleine bureaus – bij de interpretatie van testresultaten deze terughoudendheid niet altijd even sterk aanwezig is. De grote bureaus hebben vaak eigen afdelingen 'Research & Development' die de (eigen) tests onderzoeken, omdat het zo nadrukkelijk als probleem wordt ervaren: allochtone kandidaten dreigen buiten de boot te vallen in de selectieprocedure met bestaande tests en normen, terwijl maatschappelijke instanties en bedrijven juist meer allochtonen willen opnemen. Dan moet er een alternatieve methode gevonden worden om te zorgen dat de cliënt de norm haalt, bijvoorbeeld door het afnemen van andere tests.

Bij de kleine bureaus is er minder aandacht voor en daar maakt men zich grote zorgen over. De kwaliteitscontrole ontbreekt en een testscore wordt zowel door testgebruiker als opdrachtgever als een heilige wet gezien en niet of nauwelijks genuanceerd.

In het Onderwijsveld wordt er door verschillende mensen getwijfeld aan de mate van terughoudendheid van anderen in hun werkgebied. Regio speelt hier waarschijnlijk een rol (buiten de Randstand is men mogelijkterwijs minder doordrongen van de pro-

¹² Zie bijlage 3 voor de complete vragen.

blematiek), ervaring (jonge, onervaren testgebruikers zijn er minder in geïnteresseerd), en soms zijn andere – financiële – factoren belangrijker dan zorgvuldig testgebruik.

Men ervaart wel dat men in de Onderwereld bereid is zich bij te scholen en heeft daarom goede hoop.

In de praktijk van de Geestelijke Gezondheidszorg zijn de meningen verdeeld. Sommigen zijn overtuigd van de integriteit van hun collega's, zij zien in het algemeen een groot verantwoordelijkheidsgevoel bij (test)-psychologen. Daarnaast is het zo dat in dit werkveld oordeelsvorming altijd een 'group effort' is: er vindt discussie plaats met collega's uit verschillende disciplines, zeker bij discrepantie van onderzoeksgegevens. Ook worden er vaak door meer dan één beoordelaar onderzoeken gedaan. Om deze reden zijn de consequenties niet zo groot als er iemand tussen zit die niet terughoudend is, de testgebruiker heeft niet het laatste woord over de onderzoeksgegevens.

Overigens is niet iedereen overtuigd van deze zorgvuldigheid en lijkt er nog veel te worden afgegaan op standaardtests en –scores. Ook hier speelt ervaring een rol, maar dan omgekeerd: "jonge mensen worden in elk geval wel grondig getraind in diagnostiek. Dat werkt ook goed uit voor allochtonen. Als ik daarente-

gen kijk naar oudere collega's, dan vraag ik me wel eens af of zij stilstaan bij dit onderwerp. Wat namelijk gebeurt, is dat de oudere collega's niet meer zelf het diagnostisch onderzoek doen, maar dat laten afnemen door een assistent en dan hun oordeel baseren op scores, zonder dat zij de cliënt zelf gezien hebben. En vervolgens concluderen ze maar wat. Ook bij autochtonen".

Bijscholing wordt hier ook dringend gewenst geacht.

Inzet van specifieke deskundigen

Er wordt niet met specifieke deskundigen gewerkt bij het testen van allochtonen. Niet bij de instellingen en bedrijven waar men zelf werkt, noch dat men weet van heeft elders. Er zijn ook zeer weinig deskundigen om op terug te vallen.

De wenselijkheid daarvan wordt ook ontkend. Iedereen met voldoende achtergrond in diagnostiek (qua opleiding en ervaring) zou dit werk moeten kunnen doen, bij voorkeur met aanvullende cursussen toegespitst op deze problematiek. Wanneer daarnaast allochtone collega's die bekend zijn met de tests geconsulteerd worden, en er bij de testconstructie en het opstellen van richtlijnen gewerkt wordt met multi-etnische teams, vallen veel problemen te onder-
vangen.

Conclusie

Voor een zo goed mogelijke interpretatie van testresultaten is het van groot belang dat de nodige terughoudendheid wordt betracht. Vaak gebeurt dat door met collega's (zo mogelijk met een allochtone achtergrond) te overleggen om het eigen oordeel zo goed mogelijk te toetsen. Inzet van specifieke deskundigen komt niet voor en wordt ook niet als noodzakelijk beschouwd, omdat er algemeen van wordt uitgegaan dat men bij voldoende kennis en vaardigheden op het gebied van de psychodiagnostiek in staat moet zijn om tot een goede oordeelsvorming te komen. Daarnaast wordt getracht om verbetering van de oordeelsvorming te bereiken door niet alleen uit te gaan van testresultaten, maar ook uitdrukkelijk gebruik te maken van bevindingen uit andere informatiebronnen (gedragsobservatie, interviews, biografische gegevens etc.).

Groot probleem is in dit verband wel het ontbreken van kwaliteitscontrole, waardoor de zorgvuldigheid in de oordeelsvorming nog wel eens te wensen over laat. Met name op het terrein van Werving & Selectie (vooral binnen de kleinere bureaus) en binnen het Onderwijs wordt niet altijd de noodzaak ingezien om terughoudendheid te betrachten. Redenen daarvoor zijn gelegen in factoren als ervaring of financiële en maatschappelijke druk. Dit leidt er in veel gevallen toe dat afgegaan wordt op standaardresultaten en dat nauwelijks oog bestaat voor de specifieke situatie. In de Gezondheidszorg is vaker sprake van overleg met collega's, maar ook hier wordt in veel gevallen op een standaardwijze gewerkt. Tegen deze achtergrond wordt het met name gewenst geacht om tot een verdere bijscholing te komen.

VII CONCLUSIE

De samenstellers van dit rapport zijn, mede op basis van de interviews, van oordeel dat op de vraag wat er de afgelopen 10 jaar, sinds het verschijnen van het rapport van de commissie-Hofstee, gebeurd is, een positief en een negatief antwoord mogelijk is. Laat ons beginnen met het positieve antwoord.

Er zijn de laatste jaren verschillende proefschriften verdedigd (Hessels, Te Nijenhuis, Van den Berg, Van Leest, Helms-Lorenz) waarin nieuwe tests onderzocht zijn die expliciet ontwikkeld zijn voor gebruik in een multiculturele samenleving. Deze proefschriften hebben waardevolle informatie opgeleverd over hoe goede tests eruit moeten zien.

Er is echter ook een negatief antwoord mogelijk. Wie vanuit een praktisch perspectief kijkt naar welke tests gevalideerd zijn voor multiculturele diagnostiek kan niet anders dan tot de conclusie komen dat er nog niet veel tests zijn. Op het terrein van persoonlijkheid is er nog steeds weinig. Wie veel met bestaande tests werkt (zoals de twintig die door de commissie-Hofstee onderzocht zijn), heeft waarschijnlijk nog minder goede woorden over om de ontwikkelingen van de afgelopen 10 jaar te beschrijven. De bestaande tests zijn nog nauwelijks aan partijdigheidsonderzoek onderworpen. De tekortkomingen van deze tests zijn nog steeds niet opgeheven. Evenmin zijn er protocollen of richtlijnen ontwikkeld die aangeven wat te doen als nauwelijks tests beschikbaar zijn voor een bepaalde cliënt. Drie grote Amerikaanse organisaties op het gebied van psychologische diagnostiek (de American Educational Research Association, de American Psychological Association en de National Council on Measurement in Education) brengen richtlijnen voor testontwikkeling en -gebruik uit (de zgn. "Standards"). De laatste versie dateert van 1999 en het is opmerkelijk hoeveel aandacht in deze richtlijnen bestaat voor de multiculturele samenstelling van de Amerikaanse testpopulatie. Het behoeft geen betoog dat Nederlandse richtlijnen (mogelijk een aangepaste versie van de Amerikaanse) een belangrijke rol kunnen spelen in onderwijs en praktijk.

Psychologen zijn nog steeds niet in staat dezelfde standaarden aan te leggen bij het testen van autochtonen en allochtonen. Deze standaarden (zoals betrouwbaarheid en validiteit) worden binnen de psychologie als discipline breed gedragen (hoewel niet altijd even consistent toegepast). In het geval van het testen van allochtonen zijn zulke standaarden niet te hanteren omdat nog maar weinig tests onderzocht zijn. Er is nog veel werk aan de winkel.

BIJLAGE

- 1 SAMENVATTING INTERVIEWS
- 2 PERSONALIA VAN DE GEÏNTERVIEWDE DESKUNDIGEN
- 3 VRAGENLIJST 'DESKUNDIGEN OVER HET TESTEN VAN ETNISCHE MINDERHEDEN'
- 4 CONCLUSIES EN AANBEVELINGEN UIT HET RAPPORT 'TOEPASBAARHEID VAN PSYCHOLOGISCHE TESTS BIJ ALLOCHTONEN' VAN DE COMMISSIE-HOFSTEE
- 5 DE TESTS VAN HET RAPPORT 'TOEPASBAARHEID VAN PSYCHOLOGISCHE TESTS BIJ ALLOCHTONEN' VAN DE COMMISSIE-HOFSTEE
- 6 OVERZICHT VAN DE IN DE INTERVIEWS GENOEMDE TESTS
- 7 LITERATUUR
- 8 SUGGESTIES, PER WERKVELD EN ALGEMEEN

BIJLAGE 1 SAMENVATTING INTERVIEWS

1

Paul Abell, EGA HRMConsult

Bekendheid richtlijnen commissie Hofstee

‘Ik ben betrokken geweest bij de expertmeeting die voorafging aan het rapport.’

Genoemde richtlijnen commissie Hofstee

‘Geen van de tests is zonder meer te gebruiken.’

Geldigheid conclusies commissie Hofstee

‘Er is in de tussentijd wel meer gebeurd. Wat culturele bias betreft is bij mijn weten geen van de bestaande tests bijgesteld. Er kan dus nog steeds sprake zijn van het niet-begrijpen van items bij met name de persoonlijkheidsvragenlijsten waar veel gebruik wordt gemaakt van idioom (spreekwoorden, staande uitdrukkingen en dergelijke). Je kan je afvragen wat je dan meet.’

‘Er is veel onderzoek gedaan naar differentiële predictie. Als er al sprake is van differentiële predictie blijkt overigens dat er eerder sprake is van overvoorspelling dan ondervoorspelling van allochtone kandidaten. Het onderzoek naar itembias is meer een academische kwestie gebleken. Het effect van het weghalen van gebiasde items is te verwaarlozen.’

‘Ik denk dat sociaal-economische klasse veel meer bepalend is dan etnische afkomst. De allochtone bevolkingsgroep is meer vertegenwoordigd in de lagere sociaal-economische klassen; personen uit de lagere sociaal-economische klassen scoren, vooral door het opleidingsniveau, in het algemeen lager op tests. Ik verwacht echter de komende 10 jaar een verbetering in de doorstroom van allochtonen naar allerlei functies, veroorzaakt door de krappe arbeidsmarkt en het beter wordend opleidingsniveau van allochtonen. Je hebt nu veel meer 2e en 3e generatie allochtonen die wat opleiding betreft niet onderdoen voor autochtone leeftijdsgenoten. Ik denk dat uiteindelijk de verschillen autochtoon-allochtoon grotendeels zullen verdwijnen. Er komen ook allochtonen op hogere posities. Ik zie vooraansnog geen tweedeling ontstaan met een etnisch subproletariaat zoals wel wordt gezegd, er ontstaan hier geen getto's. Ik heb wat dat betreft goede hoop.’

‘Bij onderzoek naar de voorspellende waarde van tests is ook criterium bias belangrijk: door de positie van allochtone werknemers – als enige vertegenwoordiger van een groep – wordt er meer op ze gelet, moeten zij zich steeds weer bewijzen. Hierdoor wordt het oordeel over het functioneren al vertekend. Daarnaast begaan beoordelaars ook vaak een attributiefout: als het fout gaat wijt men dat aan de groep waaruit iemand afkomstig is, en als het goed gaat ligt dat aan externe omstandigheden als “geluk hebben”. Een andere factor is testgewenning. Kandidaten van allochtone afkomst kunnen bijvoorbeeld minder ervaring hebben met multiple choice vragen of vragen beantwoorden onder tijdsdruk. Mijn indruk is dat allochtone kandidaten vaak een verkeerde strategie hanteren bij het oplossen van vragen onder tijdsdruk: zij letten meer op de kwaliteit dan op kwantiteit van hun antwoorden.’

Genoemd uitgevoerd onderzoek

‘Uit de onderzoeken van Van Leest en Te Nijenhuis is gebleken dat tests wel te gebruiken zijn, met inachtneming van de genoemde beperkingen.’

‘Er zijn ook videotests ontwikkeld die gebruikt worden bij onderzoek naar sociale competentie, deze zijn goed te gebruiken want er zit geen geschreven-taalcomponent in. Het gaat daarbij om het beoordelen van situaties met sociale interactie, als indicatie voor hoe men in zo'n situatie zelf zou handelen. En dan blijkt dat de verschillen tussen autochtone en allochtone kandidaten aanmerkelijk afnemen.’

Voorstellen voor verder onderzoek

‘Er zou onderzoek gedaan kunnen worden naar criteriumbias: het feit dat iemand afkomstig is uit een andere cultuur kan de beoordeling van praktijkresultaten beïnvloeden. Beoordelaars kunnen hierop getraind worden, of er zouden speciale beoordelingsinstrumenten (b.v. gedragsschalen) ingevoerd kunnen worden.’

‘Er gaat veel aandacht uit naar hoe allochtone kandidaten binnengehaald kunnen worden, maar er zou vooral ook gekeken moeten worden hoe allochtone werknemers binnen gehouden kunnen worden.’

Waarom heeft dit soort onderzoek niet plaatsgevonden?

‘Dat er zo weinig onderzoek plaatsvindt naar de voorspellende waarde heeft vooral te maken met problemen van onderzoekstechnische aard: te kleine en te diverse allochtone groepen, en contaminerende factoren zoals criteriumbias en sociaal-economische klasse. Deze problemen hebben als praktische consequentie dat onderzoek veel tijd vergt, veel werk en dus duur is. Verder zal een selectiebureau niet blij zijn als uit het onderzoek naar voren komt dat de gehanteerde tests niet deugen.’

Genoemde gereviseerde of nieuwe tests

‘De Multi-Culturele Test MBO (MCT-M), ontwikkeld door NOA, maar deze moet zijn waarde nog bewijzen. De GATB, en de videotest.’

Terughoudendheid in interpretatie

‘Ik ga goed na of alle zaken goed begrepen zijn. In een testbatterij die wij speciaal ontwikkeld hebben voor mensen van allochtone afkomst die werkzaam zijn in de Sociale Werkvoorziening, wordt gebruik gemaakt van beeldmateriaal en gesproken woord in plaats van geschreven tekst. Bij de GATB-afname geef ik uitgebreide instructie, laat veel oefenen, geef voorbeelden, leg uit wat dingen betekenen, doe dingen voor.’

‘De testpsycholoog moet over enige multiculturele competentie beschikken, en over inlevingsvermogen in een andere cultuur.’

Terughoudendheid in interpretatie bij anderen

‘Van de grote bureaus weet ik het niet, maar vaak is er geen welwillende houding bij testpsychologen. Er wordt niet gekeken naar de testvoorspellende waarde, er is geen tijd voor en men durft niet, want men loopt het risico in de eigen vingers te snijden. In de praktijk is het weliswaar erg moeilijk om tests goed te valideren en beoordelen op bias (onbetrouwbaarheid, de tijd, de spreiding, te kleine groepen); kortom, je kunt ze niet alles verwijten, maar toch vind ik dat veel psychologen voetstoots aannemen dat hun testmateriaal ook betrouwbaar en valide is als het gaat om allochtone kandidaten.’

‘In de praktijk zie ik vaak dat kleine bureaus eigen tests gebruiken, waarvan de kwaliteit zeer discutabel is.’

‘Voor veel van onze opdrachtgevers is het op een betrouwbare wijze testen van allochtonen wel een issue. Overheden b.v. bellen regelmatig en willen dan cultuurvrije tests. Zij vragen eigenlijk onmogelijke dingen. Dat betekent dat zij of terugvallen op de grote bureaus die werken met de bestaande tests – met de eerder genoemde beperkingen - of bij die kleine bureautjes terecht komen, die van alles beloven en dan maar wat aan vogelen.’

Inzet van specifieke deskundigen

‘Wij testen veel mensen van allochtone afkomst, we zijn zelf deskundig geworden. Ik laat de resultaten van praktijksimulaties meestal zwaarder wegen dan de resultaten van tests.’

‘Bij mijn weten schakelt geen enkel bureau specifieke deskundigen in’.

Ernst van Beemen, GITP International BV

Binnen GITP heeft Ernst van Beemen met enkele collega's knelpunten en mogelijke oplossingen ("best practices") voor interculturele assessments geïnventariseerd. Ze hebben richtlijnen (tips en adviezen) geformuleerd voor hun collega's. Concept visie en beleid zijn geformuleerd. R&D vragen komen naar voren.

Inhoud, gebruik en bekendheid van de richtlijnen en conclusies van commissie Hofstee.

‘Er had veel meer benadrukt kunnen worden dat er meer is dan ethnocentrisme alleen waarop gelet moet worden bij het screenen, construeren en gebruiken van tests. Taalgebruik, zoals dubbele ontkenningen in een zin of antwoord (Gordon, NEO PI-R), staande uitdrukkingen, gezegdes, beeldspraak, niet eenduidige formuleringen, verschillen in betekenissen van woorden, bijvoorbeeld gebruikt in Suriname, op de Antillen of in Nederland. Het duurt heel lang voor dat men een vreemde taal in nuances beheerst en er dus minimale ruis bestaat bij het afnemen van een test die taalafhankelijk is. Allochtonen die goed Nederlands spreken, zijn niet altijd geheel vernederlandst. Dit wordt vaak over het hoofd gezien. Bij het beantwoorden van testvragen luistert dit volgens erg nauw.’

‘De richtlijnen en conclusies van de commissie Hofstee waren destijds dus beperkt. Ook lag de focus hoofdzakelijk op psychometrische aspecten. De test als ethnocentristisch middel werd in dit licht nooit ter discussie gesteld. Een vrij enge visie op diagnostiek, ook nu, klinkt bij het NIP door, uitsluitend op basis van psychologische vragenlijsten. Het lijkt wel of de psycholoog zich krampachtig vasthoudt aan een test en zijn professionaliteit en verantwoordelijkheid niet breder durft te zien.’

‘Sinds de richtlijnen van die commissie is er veel meer ervaringen, inzichten en feiten naar voren gekomen. Ik geloof trouwens, dat die richtlijnen niet breed bekend en verspreid zijn bij testconstructeurs en gebruikers. Maar ook is er nog te weinig onderzoek gedaan in breder verband samen met allochtonen en vooral allochtone psychologen. Incidenteel worden stappen gemaakt in de goede richting. Maar de Nederlandse psychologenwereld laat zich liever zien met een academische polemiek (zie *de Psycholoog*) vanuit ivoren torens dan dat de vragen uit praktijk als uitgangspunt worden genomen. Er is sinds de commissie Hofstee eigenlijk te weinig gedaan en kostbare tijd ging verloren, want het probleem van onrechtvaardige beoordeling van allochtonen is alleen maar schrijnender geworden.’

‘Op basis van wat nu bekend is zouden richtlijnen voor constructie, toepassing, gebruik en interpretatie van testen in dit licht nogmaals goed op een rijtje kunnen en moeten worden gezet. Destijds was het iets van een opwelling of politieke geste, want er is weinig follow-up aan gegeven. Wat wij bij GITP gedaan hebben zou verder kunnen worden uitgebreid en verbreed.’

Thema's voor verder onderzoek

‘In de eerste plaats zou men het begrip en idee van intercultureel assessment inderdaad moeten verbreden met assessment simulaties en interviews. Maar ook het testgedrag van allochtone deelnemers is een belangrijk onderwerp van aandacht en onderzoek. Van belang blijft natuurlijk ook de aandacht voor de verschillen in ontwikkeling van capaciteiten, persoonlijkheid, gedrag, invloed van cultuur en voorspellende competenties voor succes in een Nederlandse werksituatie (bijvoorbeeld het onderzoek met Multiculturele Persoonlijkheden Vragenlijst (MPV) door Van der Zee & Van Oudenhoven vanuit de RUG).’

‘Het intercultureel bewustzijn bij psychologen is een belangrijk aspect in het geheel. Bij de meeste Nederlandse psychologen lijkt niet erg sterk ontwikkeld. Dat heeft een enorme invloed op interpretatie van indrukken uit diagnostische middelen. Psychologen en andere beoordelaars zouden zich eerst veel meer zelf bewust moeten worden van eigen culturele programmering en de invloed daarvan op beoordeling van mensen uit een andere cultuur. Vervolgens zouden Nederlandse psychologen zich veel meer moeten verdiepen in de invloed van andere culturen op interactie, gedrag, waarden, houding, ontwikkeling van persoonlijkheid en capaciteiten. Allochtone psychologen kunnen daarbij een brugfunctie vervullen.’

‘De invloed van eigen culturele programmering en een beperkt intercultureel bewustzijn op het verwerven en interpreteren van assessmentindrukken (inclusief testen) is waarschijnlijk groter dan we denken. Het is essentieel dat psychologen bijvoorbeeld uitkomsten van persoonlijkheidsvragenlijsten kunnen relativiseren tegen de culturele achtergrond van een kandidaat en dit in juiste proporties kunnen gebruiken om gedrag in een Nederlandse setting te voorspellen. Voorwaar een hele opgave.’

‘Verder zouden alle nu bekende factoren die van invloed zijn op assessmentgedrag en -resultaten van allochtonen op systematische manier en in samenhang in kaart moeten worden gebracht en onderzocht. Samenwerking met allochtone deelnemers en psychologen is hierbij een must.’

‘Zo zijn er veel meer onderzoeksthema's die op een rijtje kunnen worden gezet en die in een

duidelijk gestructureerd onderzoeks-programma breed moeten worden aangepakt. Overheid en universiteiten kunnen hiertoe samen initiatief nemen. Ook het NIP kan hierbij een belangrijke rol spelen.'

Nieuwe tests en criteria voor testkeuze

'Bij de samenstelling van een assessmentprogramma en de beoordeling van kwaliteiten moeten de competenties, die nodig zijn om effectief in een Nederlandse context te kunnen functioneren steeds als referentiekader centraal blijven staan. Daar waar we weten of vermoeden dat we problemen krijgen bij het beoordelen van specifieke competenties, moeten we vooraf maatregelen nemen om die problemen zoveel mogelijk te voorkomen.'

'Gebruik testen met een duidelijk, eenduidig en eenvoudig taalgebruik. Gebruik algemene brede band persoonlijkheidsvragenlijsten (bijv. GEPPS) en geen specifiek gerichte (bijv. NPV). Gebruik bij het meten van intelligentie een breed scala van soorten testen, die zowel 'fluid' als 'crystalized intelligence' meten, en focus daarbij op de G-factor. Gebruik naast speed tests ook of liever power tests. Maak waar mogelijk ook gebruik van de Multiculturele Capaciteiten Test (MCT-M of -H) om de hierbij behorende aanvullende hulpmiddelen en referentiegroepen te kunnen gebruiken.'

'Maak naast tests veel gebruik van goed voorbereide en gestructureerde interviews, gericht op competenties (eerdere ervaringen) en cultuurverschillen. Benut alle indrukken uit assessment-simulaties (rollenspelen), ook als aanvulling op andere assessment onderdelen. Bij een 'Fact Finding' krijgt men al snel een indruk van hoe probleemanalyse en oordeelsvorming verlopen. Deze indicatie voor informatie- verwerking kan gelegd worden naast de indrukken uit klassieke intelligentietests. Gedrag in rollenspelen (leidinggevend, adviserend, klantgericht, commercieel, etc.) geeft vaak wel een duidelijk cultuurverschil aan waarmee bij interpretatie rekening moet worden gehouden. Maak bij simulaties vooral gebruik van voor de vraagstelling herkenbare en relevante situaties. Maak geen gebruik van postbakken bij mensen die de Nederlandse taal te weinig beheersen. Lezen, inzicht en beantwoorden kosten tijd en energie en leveren door mogelijk minder begrip van de situatie(s) flinke en ongewenste ruis op.'

'Er zijn bestaande tests onderzocht op bruikbaarheid bij allochtonen (o.a. Te Nijenhuis en Van Leest). Ook zijn er nieuwe tests ontwikkeld, zoals de Multiculturele Capaciteiten Test (MCT - M/H) en Multiculturele Persoonlijkheids Vragenlijst (MPV). Maar er moet nog veel meer en systematischer onderzocht en ontwikkeld worden om psychologische tests, assessment-simulaties en interviews weer op een vergelijkbaar niveau voorspellende waarde te geven en te laten behouden, die ze hebben in de toepassing bij autochtonen.'

'Tenslotte mogen we niet blind zijn voor de ontwikkeling die noodzakelijk is om ook in een multiculturele samenleving als beoordelende en adviserende psychologen professioneel en effectief te kunnen blijven functioneren. Een congres, als dit, is dan ook een goed zaak, als het maar niet te veel gericht is op terugkijken wat er niet is gebeurd en vooral gericht is op een slagvaardige, brede en effectieve aanpak in de toekomst.'

Winfried Deijmann, Horizon Bunnik

Bekendheid richtlijnen commissie Hofstee

'Er is mijns inziens niets of weinig te verwachten van tests of richtlijnen. Zij gaan uit van een te beperkte invalshoek. De dialoog en het begeleiden van mensen zijn veel belangrijker dan het sec testen van mensen onder bepaalde richtlijnen.'

'En als je al werkt met tests, dan moeten die tenminste gemaakt worden door mensen die zich van binnenuit kunnen inleven in de positie van allochtonen.'

Bekendheid richtlijnen commissie Hofstee bij testconstructeurs/-gebruikers/getesten

'Men doet vermoedelijk wel pogingen, maar men heeft niet de juiste bril op om tot een goede integratie te komen. En met 'bril' bedoel ik persoonlijke ervaringen vanuit - en bekendheid met de cultuur van de testpersoon.'

Criteria voor testkeuze

'Ik hou niet expliciet rekening met de richtlijnen van de commissie want ik gebruik geen reguliere tests. Ik doe dit naar eigen inzicht. Je moet vanuit een andere vorm dit issue benaderen, via dialoog en begeleiding.'

'Anderen proberen er wel rekening mee te houden is mijn indruk.'

'Sommige collega's in huis werken wel met tests. Omdat de opdrachtgever dat wil b.v. Er moet nl. ook een ontvankelijkheid zijn voor onze boodschap, willen we op deze manier kunnen werken. Maar consultants bij ons kunnen in principe werken op hun eigen manier. Wij zijn een lerende organisatie.'

Genoemd uitgevoerd onderzoek

'De onderzoeken zoals zij gedaan worden, komen voort uit een te beperkt paradigma. De vragen die je stelt, zul je wel beantwoord krijgen door dit onderzoek, maar het gaat mij om de vragen die niet gesteld worden. De benadering van Leo van der Burg, in zijn boek "Doen waar je goed in bent, iedereen deugt, maar waarvoor", spreekt mij veel meer aan. Het gaat daarin om een veel bredere benadering waarin de mens centraal staat en verder wordt gekeken dan alleen gedragskenmerken. Je moet leren de juiste vragen op het juiste moment te stellen, een dialoog met mensen aangaan om zo het hele individu in beeld te krijgen. Een totaalplaatje dat ook zaken bevat die niet meetbaar zijn, zoals iemands persoonlijke zingeving, mogelijkheden en beperkingen qua talent en karakter. De onderzoeken die gedaan zijn, zijn in mijn ogen dus niet belangrijk, wat koop je er voor? Mensen ontleen er schijnzekerheden aan.'

Voorstel voor verder onderzoek

'Onderzoek zoals dat van Leo van der Burg: "Doen waar je goed in bent, iedereen deugt, maar waarvoor" vanuit een holistisch en consistent waardenrijk mensbeeld. Die stelt de culturele achtergrond centraal en maakt moeilijker grijpbare factoren zoals talent en karakter inzichtelijk in hun samenhang.'

Waarom heeft dit soort onderzoek niet plaatsgevonden?

'Onderzoek vergelijkbaar met het onderzoek van Leo van der Burg, wordt tegengehouden door de gevestigde orde, omdat het daarbij gaat om zaken die niet meetbaar zijn en daarom als niet wetenschappelijk worden afgedaan. Het huidige wetenschappelijke dogma is de grootste rem op verdere ontwikkeling in dit gebied.'

Genoemde gereviseerde of nieuwe tests

'De tests zijn niet veranderd, want die zijn gebaseerd op onderzoek van 10-20 jaar geleden en de theoretische uitgangspunten van het behaviorisme. De huidige tests en richtlijnen zijn vanuit het positivisme, het meetbare, opgesteld i.p.v. uitgaande van de culturele achtergrond van mensen. Onderzoeken worden vanuit beperkte eenzijdige paradigma's gedaan en vinden dus beperkte gegevens mijns inziens. Leo van der Burg geeft een nieuw kader. Ik gebruik eigenlijk geen tests meer, ik beoordeel mensen op mijn eigen manier, met eigen vragen, formats, en kaders. Een test is hooguit een fundament om in gesprek te komen. Je moet veel breder kijken. Ik heb de dialoog en het begeleiden in de praktijk als mijn benadering gekozen. Dat gebeurt op basis van gelijkwaardigheid. Het laat mensen hun eigen oplossingen. Ook transparantie is een belangrijk punt in mijn benadering. Mensen moeten worden geïnformeerd over wat de normen zijn van waaruit de test is ontstaan en wat de achtergrond van een test is. Zingeving is ook een belangrijk onderdeel. Mensen vragen zich steeds meer af: waarvoor werk ik bij dit bedrijf? Zij hebben geen behoefte aan foefjes uit de (Amerikaanse) boekenkast, maar aan een gesprek, begeleiding.'

Nieuwe richtlijnen

'Zulke richtlijnen vind ik niet gewenst. Het gaat, zoals ik eerder al zei, uit van een te beperkt denkkader. Ze zijn goed bedoeld en voorkomen conflicten, maar dringen niet door tot de essentie. Je moet mensen leren de juiste vragen te stellen in een dialoog met mensen. Men moet buiten het testbare in gesprek verder komen.'

Terughoudendheid in interpretatie bij anderen

'Men is vergeten dat zoiets als objectieve interpretatie NIET bestaat, ieder interpretatie is altijd subjectief, persoons- en theoriegebonden. Uit onmacht of wetenschappelijke arrogantie redeneert men in het algemeen in de trant van: het is niet anders, er zit nu eenmaal vertekening in die test. En dat is een zwaktebod.'

Inzet van specifieke deskundigen

'Er dienen mensen die van binnenuit de problematiek kunnen begrijpen betrokken te worden bij het beoordelingsproces. Dat zijn mensen die in twee werelden kunnen leven, de westerse en de betreffende lands-cultuur.'

Hans van Doorn, Centrum Educatieve Dienstverlening (CED) Rotterdam

Geldigheid conclusies commissie Hofstee

'Voor een deel gelden die nog steeds. Een aantal tests is ethnocentrisch. Dat er sprake zou zijn van discriminatie, gaat me overigens te ver.'

Genoemde richtlijn commissie Hofstee

'Voorzichtig met interpretatie, m.n. bij verbale tests.'

Bekendheid richtlijnen commissie Hofstee testconstructeurs/-gebruikers/getesten

'Bij de testconstructeurs neem ik aan van wel. De mensen van de VU zoals Drenth en Bleichrodt, zijn daar erg mee bezig. Bij de testgebruikers valt het tegen. De oudere collega's zijn minder geïnformeerd.'

Criteria voor testkeuze

'Het zijn niet zo zeer deze richtlijnen. Het is meer alle informatie die ik doorneem, de studies die ik lees, de artikelen, daardoor blijf ik op de hoogte. Dat heb je dan in je achterhoofd bij het beoordelen van tests. Hoe anderen dat doen, dat weet ik niet, maar ik denk op ongeveer dezelfde manier.'

'De specifieke situatie van de te testen persoon bepaalt mijn keuze voor een bepaalde test. Voor verschillende delen van een test geldt een verschillende mate van validiteit. De verbale, performale en abstracte onderdelen zijn in afnemende mate mogelijk vertekend door ethnocentrisme. Daar hou je rekening mee. Als een leerling weinig Nederlands spreekt, kijk je meer naar de performale en abstracte onderdelen.'

Uitgevoerd onderzoek

'Het onderzoek voor de LEM, de leertest van Resing. Daaruit blijkt dat er een standaardafwijking is b.v.'

Voorstel voor verder onderzoek

'Onderzoek naar persoonlijkheidstests. De Big Five naar culturele achtergrond uitsplitsen voor adolescenten b.v.'

Genoemde gereviseerde of nieuwe tests

'In de GIVO is er waarschijnlijk rekening meegehouden. En de NDT - de opvolger van de IBO Differentiatietest, die nu uitkomt, daar is er ook wel rekening meegehouden neem ik aan. Er zullen waarschijnlijk normen voor interpretatie van allochtone scores zijn.'

Nieuwe richtlijnen

'Dat wat er ontwikkeld is, is weinig vernieuwend. B.v. de richtlijn verbale tests met een korrel zout te nemen, de conclusie dat niet-verbale tests ook ethnocentrische bias (kunnen) hebben.'

'Wat er is ontwikkeld aan richtlijnen, is wel algemeen toepasbaar.'

Wenselijkheid van en eisen aan nieuwe richtlijnen

'De richtlijnen moeten toepasbaar zijn op alle allochtonen en culturele bias dient in de handleiding vermeld te worden.'

'Er moet niet vergeten worden de richtlijnen ook uit te dragen. Dat is erg belangrijk.'

Terughoudendheid in interpretatie

'Ik verzamel zo veel mogelijk informatie naast de tests.'

'Ik overleg met één of meer collega's, al dan niet van allochtone afkomst.'

'Het is ook goed om inzicht in de beheersing van de eigen taal van de geteste te hebben. Dan raadpleeg ik wel eens een collega die de taal beheerst.'

'Ik laat me leiden door ervaring bij het scoren van tests.'

'Bij rapportages niet negatief formuleren, ook al heb je je twijfels moet je toch positief formuleren. Je moet geen pedagogisch negativisme aankweken bij de docenten.' 'Bij verbale gedeelten ben ik terughoudender dan bij performale gedeelten, want die zijn meer vertekend.'

Terughoudendheid in interpretatie bij anderen

'Ik zie dat er op vrij grote schaal gebeunhaasd wordt. De RVC die tienduizenden leerlingen test, verwijst zoveel mogelijk leerlingen naar het zorgonderwijs, want dan krijgen ze f 5.000 per leerling. Terwijl ze eigenlijk helemaal niet mogen testen.'

'Bij oudere collega's zie je wel laksheid en lacunes in kennis. Zij zouden eigenlijk bijgeschoold moeten worden, want de aantallen allochtonen groeien.'

In toevoeging op wat besproken werd

'Het is ook belangrijk om woorden uit de taal van te testen persoon te kennen. Je moet wel iets weten van die culturen, al is het maar een beetje. Zo stel je ze op hun gemak.'

Arne Evers, Universiteit van Amsterdam, Arbeids- & Organisationspsychologie

Bekendheid richtlijnen commissie Hofstee

'Ik ben betrokken geweest bij het opstellen van de richtlijnen destijds. En ik ben lid van de Commissie TestAangelegenheden (COTAN). De COTAN-richtlijnen m.b.t. testscreening zijn aangepast door de richtlijnen van de commissie.'

Genoemde richtlijnen

'Wees voorzichtig met het testen van allochtonen i.v.m. taalvaardigheid en culturele misperceptie van items, ga niet klakkeloos uit van de validiteit van de test, en verifieer de interpretatie met aanvullende informatie.'

Bekendheid richtlijnen commissie Hofstee bij testconstructeurs/-gebruikers/getesten

'Bij de testconstructeurs wel, bij de kandidaten niet. Bij nieuwe tests letten de testconstructeurs op ethnocentrisme en als testbeoordelaar kom ik weinig tegen. Bij de testgebruikers heb ik mijn twijfels: het rapport is naar mijn idee snel vergeten bij de nieuwe lichten teststers ondanks de tam-tam destijds. Het rapport is wel mede aanleiding geweest voor de promotieplaatsen van Te Nijenhuis, Van den Berg en Van Leest. Het paste in de context van die tijd.'

Conclusies commissie Hofstee

'De conclusies gelden eigenlijk helemaal niet meer. Uit onderzoek blijkt dat tests over een goede voorspellende waarde beschikken en dus valide genoemd kunnen worden. Er wordt bij testconstructie gelet op ethnocentrisme en onderzoek naar itembias wijst uit dat de invloed daarvan te verwaarlozen is. Dit geldt echter vooral voor de tests die gebruikt worden in de werving en selectie, zoals capaciteitentests. Bij de persoonlijkheidstests is niets veranderd. De idiomatische teksten die daar dus nog instaan, zouden nog steeds problemen kunnen opleveren. Bij deze tests, waar het om genuanceerde persoonlijke kenmerken gaat, waarvan de waarde per cultuur kan verschillen, kun je

daarom nog tegen problemen aanlopen. Omdat items niet begrepen worden of heel anders geïnterpreteerd. Maar dat gebied uitgezonderd, denk ik dat we ons om de meeste van de conclusies niet meer zo druk hoeven te maken.'

'Van elke selectieprocedure vormt een persoonlijkheidstest ook altijd een onderdeel, naast een IQ-test en een motivatielijst, etc. Maar in principe hebben de capaciteitentests een hogere validiteit en daar zou dus een groter gewicht aan toegekend moeten worden. Of dat gebeurt is vers twee.'

Tekortkomingen richtlijnen commissie Hofstee

'Onderzoek komt er te weinig in tot uitdrukking. Het is te veel op klinisch oordeel gebaseerd.'

Genoemd uitgevoerd onderzoek

'De conclusie van het onderzoek van Te Nijenhuis is dat er weinig problemen zijn met capaciteitentests. En uit de studie van Van den Berg komt naar voren dat een taalcomponent in een test allochtonen benadeelt. Maar deze kan er toch niet uitgehaald worden, want dat tast de voorspellende waarde van de test aan.'

'Onderzoek wijst uit dat het met de benadeling van allochtonen meevalt. Uit testtechnische gegevens blijkt dat scores een goede voorspellende waarde hebben. De tests zijn dus redelijk eerlijk. De rest moet komen van opleiding, training en taalcursussen, niet van tests.'

'Veel van het goede nieuws dringt niet door, zoals de goede voorspellende waarde van tests die in de werving en selectie gebruikt worden. Uit Amerikaans meta-onderzoek komt naar voren dat het erg meevalt met testbias. Hoewel het wel zo is dat de Nederlandse situatie mogelijk anders is dan in de VS. De diverse etnische groepen daar spreken in elk geval allemaal Engels, maar b.v. de vluchtelingen hier spreken geen Nederlands. Helemaal vergelijkbaar is de situatie dus niet.'

Voorstel voor verder onderzoek

'Onderzoek naar het testen van vluchtelingen. De problematiek ligt voor die groep toch wel anders dan voor allochtonen die hier al veel langer zijn.'

Genoemde gereviseerde of nieuwe tests

'De BIT is gereviseerd sinds 1990.'

Criteria voor testkeuze

'Ik ben geen praktische testgebruiker, maar als testconstructeur hou ik er heel expliciet rekening mee bij instructies en aanwijzingen voor de geteste, om die zo eenvoudig, simpel en duidelijk mogelijk te maken, met voorbeelden b.v. Daarnaast ben ik ook COTAN-lid en beoordelaar van tests. In het testbeoordelingssysteem zit een vraag of er items zijn die mogelijk voor allochtone cliënten minder geschikt zijn, b.v. omdat zij een racistische inhoud hebben of idiomatische problemen zouden kunnen opleveren. In die zin hou ik er als testbeoordelaar rekening mee.'

Genoemde nieuwe richtlijnen

'De ITC Guidelines voor het testen van allochtonen, daarvan is in 2000 de Nederlandse vertaling verschenen. In voortgaande discussies in b.v. *De Psycholoog*¹ en andere artikelen, worden ook richtlijnen en tips gegeven, die bij elkaar genomen een goed inzicht geven. Die richtlijnen zijn, ondanks dat ze afkomstig zijn uit de werving en selectie en loopbaanbegeleiding, wel algemeen toepasbaar.'

Terughoudendheid in interpretatie bij anderen

'Er is niet veel onderzoek gedaan naar terughoudendheid bij de interpretatie.'

'In het werkveld van de werving en selectie wordt er wel veel rekening gehouden met dit issue, omdat het zo nadrukkelijk als probleem wordt ervaren: welke tests moet men gebruiken of moet men tests aanpassen. Allochtone kandidaten scoren lager, dreigen buiten de boot te vallen, terwijl men juist meer allochtonen wil aannemen. Er moeten dan alternatieve methoden gevonden worden om te zorgen dat de cliënt hun norm haalt. Of dat een norm is die je op dezelfde test lager stelt of dat je een andere test afneemt. In de praktijk moet je meestal denken aan het afnemen van tests op een eenvoudiger niveau als het gaat om capaciteitentests.'

¹ *De Psycholoog* is het tijdschrift van het NIP, dat maandelijks verschijnt. In *De Psycholoog* is regelmatig gepubliceerd over testgebruik bij allochtonen (zie de Bijlage Literatuur).

Inzet van specifieke deskundigen

‘Iedereen met een diagnostiek-aantekening zou dit werk moeten kunnen doen, bij voorkeur met aanvullende cursussen toegespitst op deze problematiek. B.v. de postdoctorale cursus van het NIP Dienstencentrum.’

David Ingleby, Universiteit Utrecht, Cross-Cultural Studies

Geldigheid conclusies commissie Hofstee

‘Voor een deel gelden die nog. Er zijn nl. nog oude tests in gebruik, die niet herzien zijn voor ethnocentrisme. Maar er zijn ook nieuwe tests ontwikkeld, zoals de MCT-M, en die bevatten geen ethnocentrisme meer, en een aantal oude tests wordt niet meer gebruikt. Het blijft echter riskant om de tests voor vergelijkend oordeel te gebruiken, omdat er niet gekeken is naar “construct validity”. Dat wil zeggen dat er niet is gekeken of je wel de juiste vragen stelt om een bepaalde eigenschap, kenmerk of klacht te meten voor andere groepen dan de Nederlandse.’

Bekendheid richtlijnen commissie Hofstee

‘Het rapport is één van de basisdocumenten over dit onderwerp. Van de 20 meeste gebruikte tests was er geen één die vrij was van een vermoeden van bias.’

Bekendheid richtlijnen commissie Hofstee bij testconstructeurs/-gebruikers/getesten

‘Op de testconstructeurs heb ik geen zicht, maar van studenten die stage lopen, hoor ik dat er in de praktijk weinig terughoudendheid wordt betracht in interpretatie van testresultaten van allochtonen. De tests worden vaak gebruikt als heilige schrift.’

Tekortkomingen richtlijnen commissie Hofstee

‘De commissie is tot zijn oordeel gekomen zonder empirisch onderzoek. Er is verder geen aandacht voor de afname van tests en de interpretatie van testcores. Tests neem je niet af in een vacuum. En er is weinig aandacht voor de klinische tests. Bij de 20 tests die gescreend zijn, zat er niet één uit de GGZ. De richtlijnen worden dan ook niet opgevat als ook geldig voor tests in het klinische veld.’
‘In plaats van over autochtonen en allochtonen te praten moet er specifiek naar verschillende groepen gekeken worden, zoals b.v. Marokkanen van de eerste generatie, de tussengeneratie en de tweede generatie, en er moet ook onderscheid gemaakt worden tussen Berbers en Arabisch sprekkenden. Dit soort dingen moet terugkomen in richtlijnen.’

‘Amerikaanse conclusies zijn niet zonder meer generaliseerbaar naar de Nederlandse situatie. Het verbaast me dat sommige Nederlandse onderzoekers vinden dat ze zomaar kan generaliseren van onderzoeken naar zwarte Amerikanen naar allochtonen in Nederland. Je moet nauwkeurig kijken naar de geschiedenis en maatschappelijke positie van een groep, de taal die ze spreken en hoe lang geleden ze geïmmigreerd zijn.’

Genoemd uitgevoerd onderzoek

‘Je hebt het boek van Bleichrodt en Van de Vijver: “Diagnostiek bij Allochtonen”. En de discussie in *De Psycholoog*² tussen Paul van Leest, Nico Bleichrodt en Remko van den Berg aan de ene kant en Evers e.a. aan de andere kant. De criterium-validiteit van bepaalde tests is goed onderzocht en het blijkt mee te vallen met de bias. Maar, nogmaals, veel conclusies zijn gebaseerd op Amerikaans onderzoek. Bovendien is er geen onderzoek gedaan naar “construct validity”.’

Waarom heeft dit soort onderzoek niet plaatsgevonden?

‘In de GGZ wordt te onkritisch gekeken naar de validiteit van instrumenten omdat men gericht is op efficiency: de kosten van de zorg moeten teruggebracht worden. Men is dus gericht op standaardisering. Maar cliënten zijn niet standaard, die verschillen, en als je daar niet op inspringt, ben je niet kostenbesparend bezig. Je verliest aan kwaliteit en accuratesse in je dienstverlening.’

² *De Psycholoog* is het tijdschrift van het NIP, dat maandelijks verschijnt. In *De Psycholoog* is regelmatig gepubliceerd over testgebruik bij allochtonen (zie de Bijlage Literatuur).

Hoe kan dit soort onderzoek bevorderd worden?

'Er moet een agressievere aanpak komen, van een alliantie van onderzoekers, praktijkmensen en cliënten. Je moet dezelfde taal spreken als de zorgmanagers en er op wijzen dat er door standaardisering sprake is van schijnefficiëntie. Adviesraden van gewicht, zoals de RVZ (Raad van Volksgezondheid en Zorg), moeten dit eigenlijk aanhangig maken bij de minister. Het voortraject, om die adviesraden te overtuigen, zou een taak voor het NIP kunnen zijn.'

Genoemde gereviseerde of nieuwe tests

'In de klinische wereld zijn er voor veel verschillende instrumenten aparte normen vastgesteld, maar de tests zijn dezelfde. De veronderstelling is dus nog steeds dat je vragen relevant zijn. In de cross-culturele psychologie spreekt men van een 'etic' of een 'emic' benadering. Etic wil zeggen universalistisch en emic vanuit de eigen denkkaders van specifieke groepen. De antropologische benadering (emic) in de psychologie ontbreekt grotendeels.'

Wenselijkheid van en eisen aan richtlijnen

'Richtlijnen zijn wenselijk maar pas na verder onderzoek. En dan m.n. met betrekking tot het afnemen van tests. De omstandigheden waaronder tests worden afgenomen beïnvloeden de resultaten. Dat blijkt b.v. uit een onderzoek bij gettokinderen. In een spelsituatie presteren ze veel beter dan in een concurrentiesituatie. Het feit dat er op hen wordt neergekeken speelt een rol. Allochtonen hier hebben vaak geen vertrouwen in Nederlandse scholen, ziekenhuizen, en bedrijven. Dat beïnvloedt waarschijnlijk de antwoorden. Je moet proberen hun vertrouwen winnen en als dat niet lukt, moet je besluiten tot een andere manier van beoordelen.'

Terughoudendheid in interpretatie bij anderen

'Men is te "easy going" en men heeft te weinig tijd. In Nederland is er meer bewustzijn dan in b.v. Engeland, men is zich bewust van de gevaren en valkuilen, maar toch wordt er niets aan gedaan. Er bestaat wat dat betreft een kloof. Men redeneert: we hebben niets anders, niets is perfect. Het blijft bij een theoretische aarzeling.'

'Het is ook een politieke kwestie: hoe open sta je voor andere normen en waarden. De psychologie houdt momenteel de Nederlandse cultuur in stand in de huidige testpraktijk. De criteria worden niet ter discussie gesteld: wat vinden wij een goede prestatie, een geschikte persoon, gepaste omgangsvormen. Ook belangrijk is de vraag: hoeveel tijd krijgt iemand om te wennen aan Nederland. Vaak zijn het nieuwkomers. Als je iemand zoekt die morgen meteen moet kunnen presteren, dan neem je geen nieuwkomer, maar als je meer tijd hebt, concessies kunt doen, kun je ook kijken naar iemand, die op den duur heel goed kan worden. Mijn werkgever had gelukkig de tijd om mij te laten wennen. Ik deed in het begin heel domme dingen. Als je mij destijds een IQ-test had afgenomen, had ik heel slecht gescoord.'

Inzet van specifieke deskundigen

'Het is in elke Riagg wel gewoonte om het oordeel te vragen van allochtone collega's, de zgn. "alibi-Ali's". Dat is een noodzakelijke tussenoplossing. Op den duur moet iedereen echter in staat zijn rekening te houden met cultuurverschillen. Er moeten daarvoor meer interculturele leergangen komen, want meestal is er te weinig aandacht voor cultuurverschillen in het curriculum.'

Wim Kleijn, Centrum '45 Asielzoekerscentrum

Genoemde richtlijn commissie Hofstee

'Voorzichtigheid in interpretatie van testcores. Ik heb er een globale herinnering aan, maar ga meer uit van alle verzamelde kennis en informatie, die ik in de loop der jaren heb opgedaan middels eigen onderzoek en artikelen van andere onderzoekers.'

Bekendheid richtlijnen commissie Hofstee bij testconstructeurs/-gebruikers/getesten

'Testgebruikers zijn zich wel bewust van de problematiek, want ze hebben er heel direct mee te

maken in de communicatie met de cliënt, maar gebruiken niet per se deze richtlijnen. Het werkt voor hen zoals voor mij denk ik. Je hebt in zijn algemeenheid deze overwegingen in het achterhoofd.'

Tekortkomingen richtlijnen commissie Hofstee

'Het is onduidelijk wat het empirisch onderzoek is op grond waarvan ze opgesteld zijn. En ze zijn niet specifiek genoeg. Het zijn vaak ook 'common sense'-richtlijnen waar je door de praktijk vanzelf tegenaan loopt en dan zelf ook wel kunt bedenken.'

'Tests moeten niet alleen beoordeeld worden op kwaliteit. Er moet ook bekeken worden welke test je moet kiezen in welke situatie.'

Uitgevoerd onderzoek

'Er is gebrek aan onderzoek naar tests die gebruikt worden in de geestelijke gezondheidszorg, met name de persoonlijkheidstests.'

Voorstellen voor verder onderzoek

'Onderzoek naar het testen van vluchtelingen, klachteninventarisatie, persoonlijkheidsdiagnostiek. Als er goede tests gemaakt worden voor vluchtelingen, zullen die ook vaak goed bruikbaar zijn voor andere allochtonen. Naar mate je vroeger in het hulpverleningstraject zit, zijn de communicatieproblemen nl. groter. Dus als je het bij vluchtelingen kunt onderzoeken, dan kun je het zeker ook bij andere groepen allochtonen.'

'Er moet ook uitgezocht worden, wanneer welke test het beste werkt.'

'Onderzoek naar de SCL-90, een klachteninventarisatie. Er zijn half-illegale vertalingen in omloop en nu pas wordt er gewerkt aan een Turkse en Marokkaanse versie. Er is niet veel gedaan de laatste 10 jaar. Veel tests kun je niet gebruiken wanneer de cliënt geen Nederlands spreekt. Dan moet je andere methoden gebruiken. Je komt dan vaak niet veel verder dan een standaardanamnese langs klinische weg. Dat heeft grote nadelen.'

Waarom heeft dit soort onderzoek niet plaatsgevonden?

'Onderzoek in de geestelijke gezondheidszorg blijft achter, omdat de uitkomsten van zulk onderzoek een onduidelijk belang hebben. Er is geen direct economisch belang te identificeren zoals dat wel het geval is in de werving en selectie en de gevolgen van het uitblijven worden maar op een klein aantal punten gevoeld.'

'Daarnaast zijn er onderzoekstechnische problemen, zoals de geringe omvang van groepen en de diversiteit ervan. Het is voor testconstructeurs niet te doen om voor allochtonen/vluchtelingen tests te maken, tenzij zij daarin gespecialiseerd zijn.'

'En mensen redden zich in de praktijk met wat er is.'

'Maar het tij is gekeerd: er komt meer aandacht, de problemen komen meer op de voorgrond, er is meer geld op de markt om onderzoek naar vluchtelingen te doen.'

Nieuwe richtlijnen

'Er zou een brochure moeten komen over welk instrument wanneer te gebruiken is, met regelmatige updates, vanuit de praktijkervaringen. Zoals de huisartsen protocollen hebben omtrent klachten en wat daar verder aan te doen. Dan krijg je veel meer sturing in het gebruik van tests. Op welke gronden er nu gekozen wordt voor welke test is onduidelijk. En er is behoefte aan een centraal punt om dat te sturen, misschien zou het NIP dat kunnen doen, als dat maar niet betekent dat alleen de leden bereikt worden.'

'Het opnemen van allochtonen in onderzoeken en deze niet langer uit te sluiten op basis van een inclusiecriteria als "goed Nederlands kunnen spreken of lezen". Er zouden ethische richtlijnen voor subsidiegevers opgesteld moeten worden om toewijzing van geld voor dit soort onderzoek te ondersteunen. Net zoals onderzoek in de medische wereld waarvan vroeger om praktische redenen vrouwen buitengesloten werden, nauwelijks meer gehonoreerd worden.'

'En er moet een richtlijn komen over het standaard toepassen van de Nederlandse spreekvaardigheidstest bij vluchtelingen indien er Nederlandstalige tests worden gebruikt. Als het Nederlands te beperkt is, dan zou voorgeschreven moeten worden geen tests af te nemen, of een tolk in te schakelen.'

Terughoudendheid in interpretatie

'Je gebruikt andere informatiebronnen ter bevestiging van de test. Wanneer er discrepantie is ben je voorzichtig. Tests zijn maar een onderdeel van het hele onderzoek en m.n. nuttig voor evaluatie, het volgen van de interventie.'

'Er is in principe wel voldoende terughoudendheid te betrachten. Er zijn geen bijzondere valkuilen. Je kijkt altijd per persoon, je maakt een individuele beoordeling, autochtoon of allochtoon.'

Terughoudendheid in interpretatie bij anderen

Het zou me verbazen als men zich niet bewust zou zijn van de problematiek. Testgebruikers hebben er nl. heel direct mee te maken in de communicatie met de cliënt. En het lijkt me ook een vreemde beroepskeuze om als psycholoog te werken met allochtonen, als je daar geen rekening mee zou houden. Uit de enquête die ik een paar jaar geleden heb rond gestuurd naar collega's in dit vakgebied, blijkt ook dat tests niet klakkeloos worden toegepast of geïnterpreteerd. Ik zie een voorzichtige houding bij mensen die met allochtonen werken.'

'Bij klinische toepassingen vindt vaak discussie plaats met verschillende collega's, verschillende disciplines, zeker bij discrepantie van onderzoeksgegevens. Daarnaast worden er meer onderzoeken gedaan per persoon, door meer dan één beoordelaar. Het is een "group effort". De consequenties zijn dan ook niet zo groot als er iemand tussen zit die niet terughoudend is. Niemands macht is binnen zo'n werkverband zo groot dat hij/zij het hele oordeel naar zijn/haar hand kan zetten. Er zijn genoeg 'safe guards'.'

Wil Koning, GPI Amsterdam

Geldigheid conclusies commissie Hofstee

'Op de eerste twee vlakken, d.w.z. ethnocentrische bias en vergelijkend oordeel, gelden ze nog steeds. Met item- en testbias blijkt het mee te vallen, gezien recent onderzoek. Maar er is sindsdien niet zoveel veranderd, in de onderwijssector althans.'

'Mogelijk zijn er ook andere factoren die de verschillen in testscore veroorzaken. Misschien zijn sociaal-economische verschillen wel veel bepalender dan etnische achtergrond.'

Bekendheid richtlijnen commissie Hofstee bij testconstructeurs/-gebruikers/getesten

'De richtlijnen leven weinig. Het is wel bekend dat ze er zijn, maar er is onvoldoende publiciteit aan gegeven. Ze zouden meer gepromoot moeten worden.'

Criteria voor testkeuze

'De richtlijnen spelen voor mij niet zo zeer een expliciete rol. Studies spelen wel een rol op de achtergrond. Ik baseer mij op kennis van onderzoeksgegevens maar daarnaast ook op gezond verstand in mijn oordeel over tests. Dat komt terug in mijn keuze voor tests, en interpretatie en relativering van testscores. Ik kijk b.v. naar afkomst (o.m. stad/platteland, welke talen werden/worden gesproken, sociaal-economische en pedagogische cultuur, verblijfsduur in Nederland) en problematiek van het kind, m.a.w. naar de specifieke omstandigheden. Gelukkig beschikken wij in Nederland - in elk geval voor kinderen en jeugdigen - over goede niet-verbale intelligentietests en een zgn. leertest. Ook zijn er al geschikte taaltests.'

Tekortkomingen richtlijnen commissie Hofstee

'De richtlijnen zijn erg toegespitst op sollicitanten. Ze zouden ook gericht moeten worden op kinderen en jeugdigen, in een school- en jeugdhulpverleningsomgeving. In het onderwijs speelt nu onder meer het probleem van de indicatiestelling voor het speciaal onderwijs. Daarbij spelen intelligentiescores een belangrijke rol. Ondeskundige testkeuze en/of interpretatie van testscores kan hier verstrekken gevolgen hebben. Steeds meer niet-daartoe-deskundigen (schooldirecteuren, voorzitters onderzoekscommissies, intern begeleiders, maatschappelijk werkenden) baseren beslissingen op IQ-scores. Ze worden daartoe gestimuleerd door regelgeving van de overheid. Ik vind dit een erg zorgelijke ontwikkeling.'

Genoemd uitgevoerd onderzoek

'Het recente boek van Bleichrodt en Van de Vijver: Diagnostiek bij allochtonen. Voor mijn werkveld is m.n. het hoofdstuk daarin van de hand van Resing en Hessels van belang: Het meten van de cognitieve mogelijkheden en het schoolgedrag van allochtone kinderen. Dat is een belangrijk stuk voor psychodiagnostici, schoolbegeleiders en onderwijskundigen. Opvallende resultaten zijn: het valt mee met de testbias, het is nog altijd onduidelijkheid welke factoren precies de verschillen tussen autochtonen en allochtonen veroorzaken, verblijfsduur is b.v. een belangrijke factor. En er wordt aanbevolen zoveel mogelijk informatiebronnen naast elkaar te gebruiken, zoals de SON-R, RAKIT en de LEM-leertest. Verder is er niet zoveel gedaan de afgelopen 10 jaar.'

Voorstellen voor verder onderzoek

'Onderzoek naar de voorspellende waarde van tests. En longitudinaal onderzoek, om de waarde van genomen beslissingen te kunnen toetsen en hoe relevant het gebruik van tests daarbij is geweest. Of b.v. doorverwijzingen naar bepaalde vormen van speciaal onderwijs zinvol zijn geweest.'

'Verder onderzoek naar de bruikbaarheid van de SON-R, de LEM, de TAK, de leervorderingentests en de TVK-taaltest zou goed zijn.'

Waarom heeft dit soort onderzoek niet plaatsgevonden?

'Er is maar een kleine groep onderzoekers met interesse in dit onderwerp.'

'En het is erg duur om voor alleen het Nederlandse taalgebied onderzoek naar een bepaalde test te doen en tot een revisie te komen.'

Genoemde gereviseerde of nieuwe tests

'De SON-R 2½-7 en de CITO-toets meertaligheid bijvoorbeeld.'

Nieuwe richtlijnen

'Er is nog veel onduidelijk. Er zal eerst meer onderzoek gedaan moeten worden voordat er verdere richtlijnen kunnen worden ontwikkeld.'

Terughoudendheid in interpretatie

'Die terughoudendheid werkt door in de keuze van de test, de interpretatie en relativering van de testresultaten.'

'Het oordeel van de ouders en van de leerkracht die dagelijks met het kind omgaan, betrek ik bij de beoordeling. Die kunnen het kind observeren in allerlei situaties, ook in sociale situaties b.v.'

'Voldoende terughoudendheid wordt bemoeilijkt door het ontbreken van normen toegespitst op allochtonen. Bij sommige tests is dat een probleem. Wanneer er alleen Nederlandse normen bestaan is het twijfelachtig of de terughoudendheid die men probeert te betrachten voldoende is.'

Terughoudendheid in interpretatie bij anderen

'Buiten de grote steden in de randstad is men, begrijpelijkerwijs, soms minder ervaren in deze problematiek. Collega's van buiten de randstad maken soms onvoldoende gebruik van goede niet-verbale tests als de SON-R.'

Inzet van specifieke deskundigen

'In de grote steden test vrijwel elke diagnosticus in dit werkveld ook allochtone kinderen. Er worden niet zozeer specifieke deskundigen ingezet. Er zijn af en toe cursussen en studiedagen voor bijvoorbeeld schoolbegeleidingsdiensten om de nodige kennis op te doen. Maar dit gebeurt de laatste tijd weinig frequent. Deze conferentie is daarom een goed initiatief.'

Jan ter Laak, Universiteit Utrecht, Ontwikkelingspsychologie

Bekendheid met richtlijnen commissie Hofstee

'Ik ben betrokken geweest bij het opstellen van de richtlijnen destijds. En ik ben ook lid van de COTAN geweest.'

Conclusies commissie Hofstee

'Etnocentrische bias is met de huidige stand van zaken in de wetenschap ook makkelijk vast te stellen. Er is veel kennis bij het CITO over verschillende groepen, over hun schoolvorderingen.'

'Ook het probleem van het vergelijkend oordeel is bij de huidige onderzoekspraktijk goed te verhelpen. We weten inmiddels veel meer en kunnen dus verdergaand, specifiek psychometrisch onderzoek doen. Testvergelijking en testequivalentie zijn goed mogelijk, daar zijn wij goed in in Nederland. We hebben echter niet veel gedaan de afgelopen 10 jaar.'

'Ze hebben ook een betere kijk gekregen op testbias, ze kunnen laten zien dat hij er is.'

Bekendheid richtlijnen commissie Hofstee bij testconstructeurs/-gebruikers/getesten

'De testconstructeurs moeten wel op de hoogte zijn van de richtlijnen. De COTAN-richtlijnen met betrekking tot het opstellen van tests, zijn bijgesteld als gevolg van bedoelde richtlijnen. Bij testgebruikers zijn ze te weinig bekend. In de handleiding staat een groep aangegeven maar in de praktijk wordt de test ook voor andere groepen gebruikt. Men is zich niet bewust van de risico's.'

Richtlijnen commissie Hofstee in de dagelijkse praktijk

'Ik ben lid van de COTAN geweest en bij het beoordelen van nieuwe tests spelen deze richtlijnen dus expliciet een rol, want de COTAN-richtlijnen voor de beoordeling van tests zijn aangepast als gevolg van deze richtlijnen. En verder gebruik ik ze in het onderwijs dat ik geef aan mijn studenten.'

Genoemd uitgevoerd onderzoek

'Uit onderzoek blijkt dat een taalcomponent in een test alloctonen benadeelt. Maar deze kan er niet uitgehaald worden, want dat tast de voorspellende waarde van de test aan. Communicatie en taal spelen nl. toch een grote rol in elke functie.'

'Veel van het goede nieuws dringt niet door. Er is geen brug geslagen tussen theorie en praktijk. De onderzoeksresultaten zijn vooral een theoretische, academische discussie gebleven en hebben weinig doorwerking gehad. We hebben te weinig uitgelegd. Dat zou een mooie taak voor het CITO zijn.'

Waarom heeft dit soort onderzoek niet plaatsgevonden?

'De persoonlijke interesse van onderzoekers en onderzoekstradities op universiteiten spelen een belangrijke rol. En dit soort onderzoek heeft in de universitaire wereld geen status. Het is "testboerwerk".'

'En onderzoek naar scoreverschillen tussen autoctonen en alloctonen heeft een politieke lading. Er wordt al snel racisme en discriminatie geroepen. Je bent de boodschapper van het slechte nieuws als je wijst op die verschillen.'

Hoe kan dit soort onderzoek bevorderd worden?

'Er zou een samenwerkingsverband tussen universiteit en praktijkmensen moeten komen, om dit onderzoek ter hand te gaan nemen.'

'En je moet af van de politieke lading. Je moet er nuchter naar kijken. Want ook als de etnocentrische bias eruit is, vind je scoreverschillen, om wat voor reden dan ook. En dat moet juist onderzocht worden. Hoe dat komt. Misschien zijn sociaal-economische factoren wel een belangrijkere veroorzakende factor dan etnische afkomst b.v. Je vindt dit soort scoreverschillen nl. ook in India en de Gazastrook.

Je moet van dat etnocentrisme geen apart issue maken. Daarmee wordt het te politiek en kun je het niet meer onderzoeken. Dan krijgt het te veel een aura van discriminatie. En dat is ongefundeerd. Ik heb nog nooit gevallen van discriminatie meegemaakt op grond van testresultaten. Je zou er zeg maar op dezelfde manier naar kunnen kijken als naar de verschillen in testcores tussen mannen en vrouwen. Of hoogintelligenten, ook een groep apart, die vereisen ook een speciale aanpak.'

Aangeven van etnocentrisme bij nieuwe tests

'Het zou een normaal onderdeel moeten zijn, maar er wordt alleen heel algemeen aangegeven dat de test voor groepen een afwijking kan hebben. Je zou een aparte rubriek moeten maken van test-

beoordeling. Je kunt daar de Standards '99 van de American Psychological Association en de American Educational Research Association (APA/AERA) goed voor gebruiken als uitgangspunt om zoiets op te stellen.'

Gereviseerde of nieuwe tests

'Het bijhouden van instrumenten is niet goed geregeld in Nederland. In b.v. Scandinavië wel.'

'De leeropdracht van de hoogleraar, de leerstoel, bepaalt of het onderwerp testen van allochtonen een regulier onderdeel is van het curriculum of niet. Als het onderwerp daar niet expliciet in genoemd wordt, dan is het de keuze van de docent of daar ruimte voor wordt gemaakt in het onderwijs.'

Wenselijkheid van en eisen aan nieuwe richtlijnen

'Het is niet wenselijk aparte normen te hanteren of andere omstandigheden, waaronder de test wordt afgenomen. Het is juist van het grootste belang de test standaard af te nemen anders krijg je onvergelykbare resultaten. Als experiment is het interessant te zien waar je testcores door beïnvloed worden, maar daar zal de oplossing van dit probleem niet vandaan komen.'

Terughoudendheid in interpretatie bij anderen

'Ze kennen het probleem, hoe ze het in de praktijk gezien het ontbreken van goede tests oplossen, weet ik niet.'

Inzet van speciale deskundigen

'Het moet een onderdeel van het werk van iedereen zijn. Alle psychologen met een diagnostiek aantekening en aanvullend daarop gerichte cursussen, moeten dit kunnen.'

Frank van Luijk, LTP

Genoemde richtlijn commissie Hofstee

'Niet-racistische inhoud.'

Geldigheid conclusies commissie Hofstee

'Uit onderzoek blijkt dat er in de regel terecht wordt gediscrimineerd: mensen die lager scoren op tests, presteren in de praktijk veelal ook slechter. De voorspellende waarde van de meeste tests is goed. Er zijn echter onderzoekstechnische problemen, die zich voordoen bij vergelijkend onderzoek, zoals de kleine groepen allochtonen, veel verschillende groepen en de gevolgen daarvan voor betrouwbaarheid, die een goed oordeel over veroorzakende factoren bemoeilijken. En daarbij is het mogelijk dat heel andere factoren de verschillen in testscore veroorzaken, zoals b.v. de sociaal-economische verschillen. Verdere contaminerende factoren die een rol kunnen spelen, zijn criteriumbias (b.v. anders beoordeeld worden op de werkplek, meer waar moeten maken) en testgewendheid (b.v. tijdsdrukstrategie). Dit alles maakt het moeilijk uit te splitsen wat waar de oorzaak van is.'

Bekendheid richtlijnen commissie Hofstee bij testconstructeurs/-gebruikers/getesten

'Ik ga ervan uit dat de testconstructeurs ermee bekend zijn. Kandidaten niet. Wanneer mensen in de werving en selectie-procedure protesteren tegen testcores, beroepen ze zich, althans is onze ervaring, nooit expliciet op de richtlijnen van de commissie.'

Genoemd uitgevoerd onderzoek

'Uit het onderzoek van Te Nijenhuis blijkt dat de gebruikte intelligentie- en capaciteitentests beschikken over een goede betrouwbaarheid en validiteit, zij leveren zijns inziens deugdelijke resultaten op, de voorspellende waarde van de gebruikte tests is goed, scoreverschillen zijn terechte verschillen: iemand die laag scoort op een test, functioneert in de praktijk ook niet goed in de functie waarvoor getest wordt. Van den Berg en Van Leest bevelen daarnaast betere instructies aan, en geven aan dat testgewendheid een rol speelt, evenals het leren ontwikkelen van een teststrategie voor het werken onder tijdsdruk.'

'De uitgevoerde onderzoeken hebben vooral bevorderd dat men zich bewust is van de problematiek. De gevoeligheid voor het onderwerp en de voorzichtigheid in toepassing van tests en interpretatie van resultaten.'

Voorstel voor verder onderzoek

'Het verband tussen test- en criteriumbias voor verschillende groepen zou nader onderzocht moeten worden.'

Waarom heeft dit soort onderzoek niet plaatsgevonden?

'De onderzoekstechnische problemen zoals ik die eerder schetste, zorgen voor obstakels, maar daarnaast is ook de mode zoals in de 90-er jaren, momenteel een beetje over. Bovendien is het zo dat de grote groei van het aantal te testen allochtonen in de werving en selectie die verwacht werd, althans bij ons, niet is gekomen. En het probleem wordt met de 2e en 3e generatie allochtonen waarschijnlijk toch minder groot, omdat die beter Nederlands spreken, meer gewend zijn aan en bekend zijn met de Nederlandse samenleving.'

'In het verleden is wel geprobeerd om met de grote werving en selectiebureaus te komen tot een samenwerking over dit onderwerp. Dit initiatief is echter gestrand, omdat een uitwisseling van informatie over gehanteerde normgroepen niet tot stand komt in de toch wat competitieve setting van de werving en selectiebureaus. De normgroepen zijn het bedrijfskapitaal en daarmee het bedrijfsgeheim van een bureau.'

Genoemde gereviseerde of nieuwe tests

'De MCT-M en de Raven.'

'Wij hebben intern een woordenlijsttest is ontwikkeld als voortest, die wordt afgenomen voorafgaand aan het assessment, om de mate van taalbeheersing van de kandidaat te kunnen vaststellen.'

Criteria voor testkeuze

'De COTAN-beoordeling van de test wordt als referentie gebruikt bij ons.'

Genoemde nieuwe richtlijnen

'Wat er is ontwikkeld aan richtlijnen, is ondanks dat het afkomstig is uit de werving en selectie en loopbaanbegeleiding, wel algemeen toepasbaar.

'Er moeten wel uniforme normen gehanteerd worden. Allochtone kandidaten doorlopen na de speciaal voor hen ontwikkelde woordenlijsttest, bij ons verder het reguliere assessmentprogramma. Als je een aparte norm gaat hanteren, dreigt het gevaar van een kneuzennorm.'

Terughoudendheid in interpretatie

'Om de nodige voorzichtigheid in je oordeel bij het testen van allochtonen in te bouwen, overleg je met collega's: wat zou jij doen in dit geval?'

Terughoudendheid in interpretatie bij anderen

'Er zijn grote verschillen. De kleine zelfstandige bureaus, daar is er weinig tijd en dus aandacht voor. Wij testen per jaar zo'n 3000-4000 mensen, wij hebben een eigen R&D-afdeling die dingen onderzoeken.'

'Er is nogal wat zelfwerkzaamheid onder leken wat betreft testafname en die zien de testscore als een heilige wet. Ook op internet kun je tegenwoordig testjes kopen. Die trend neemt toe.'

Inzet van specifieke deskundigen

'Er wordt geen gebruik gemaakt van specifieke allochtone deskundigen. Alle consultants moeten dat in hun pakket hebben. Onze consultants volgen de PBSA-opleiding³. Daarin is een aparte module over het testen van allochtonen opgenomen. Dat is niet wijd en zijd gebruik. Alleen de grotere bureaus doen dit allemaal wel.'

³ De PBSA-opleiding is een Postdoctorale Basisopleiding voor de Selectiepsychologische Adviespraktijk, die intern gegeven wordt aan een aantal grote selectie- & adviesbureaus.

In toevoeging op wat besproken werd

‘Kwantitatief vallen de problemen wat culturele bias betreft mee, neemt niet weg dat het voor het geteste individu een groot probleem is als het mis gaat.’

Bim Mungra, op persoonlijke titel

Geldigheid conclusies commissie Hofstee

‘Die gelden nog steeds op alle punten. Er is te weinig rekening gehouden met de conclusies, de testpraktijk is niet veranderd en er is veel kritiek uit allochtone hoek.’

Genoemde richtlijnen commissie Hofstee

‘Predictieve validiteit en ‘culture fair’ werken.’

Praktijkdoorwerking richtlijnen commissie Hofstee

‘Ik gebruik de richtlijnen van de commissie in de gastcolleges die ik geef aan de Leidse Universiteit.’
‘In de testpraktijk hou ik voor zover mogelijk rekening met de richtlijnen van de commissie, maar goed testmateriaal is beperkt.’

Bekendheid richtlijnen commissie Hofstee bij testconstructeurs/-gebruikers/getesten

‘De richtlijnen zijn niet doorgesijpeld. In een beperkte groep van witte en zwarte psychologen die betrokken zijn bij de allochtone groep wel. Maar verder is er niet voldoende draagvlak en PR. Het is niet opgepakt.’

Genoemd uitgevoerd onderzoek

‘Bleichrodt aan de VU doet onderzoek naar diagnostiek bij volwassen allochtonen. Te Nijenhuis heeft gekeken naar persoonlijkheidstests en intelligentie.’

‘Verder wordt er in de sociologie wel veel onderzoek gedaan naar positie en omstandigheden van allochtonen, maar niet in de psychologie.’

‘Onderzoeksresultaten dringen niet echt door in de praktijk.’

Voorstel voor verder onderzoek

‘Met name met betrekking tot intelligentie en persoonlijkheidstests moet er meer onderzoek gedaan worden.’

‘De WISC is niet voldoende aangepast aan de allochtone doelgroep en de RAKIT is ook verouderd. De normering, betrouwbaarheid en validiteit zouden eens opnieuw bekeken moeten worden.’

Genoemde gereviseerde of nieuwe tests

‘De MCT-M, ontwikkeld door NOA; en verder de SON-R, en de LEM voor Turkse en Marokkaanse kinderen. De LEM wordt helaas te weinig gebruikt, ook vanwege onhandige praktische uitvoering.’

Criteria voor testkeuze

‘De specifieke situatie van de te testen persoon bepaalt de keuze voor een bepaalde test. Wanneer het kind nog maar kort in Nederland is, niet Nederlands sprekend is, dan heeft het geen zin om het te testen. Dat is volstrekte onzin.’

Nieuwe richtlijnen

‘Nieuwe richtlijnen zouden gekoppeld moeten worden aan die van Hofstee. De ‘Standards 99’ van de APA (American Psychological Association) en de AEA (American Educational Association) zijn daarvoor een goed voorbeeld.’

‘Er moet gewerkt worden met multi-etnische teams voor het opstellen van richtlijnen.’

Terughoudendheid in interpretatie

‘Je moet een uitgebreide instructie geven en woorden uit hun eigen taal kennen om hen op hun gemak te stellen.’

'In de jaren 70, telde ik wel eens 5 punten bij de score op van kinderen van wie ik dacht "die zitten net op de grens LOM, laat ik ze een kans geven".'

Terughoudendheid in interpretatie bij anderen

'Pas afgestudeerden van de universiteit hebben te weinig in hun opleiding gehad over de problematiek bij het testen van allochtone leerlingen. Die moeten nog de nodige ervaring opdoen. En er zijn ook nog al wat witte collega's, die weinig interesse in het onderwerp hebben. Degenen die betrokken zijn bij deze problematiek, die wel. Die houden er rekening mee.'

Inzet van specifieke deskundigen

'Er is wel speciale scholing nodig om allochtone kinderen te testen.'

'Collega's uit de allochtone groepen moeten betrokken worden bij het beoordelingsproces.'

In toevoeging op wat besproken werd

'Ook in de testconstructie en het opstellen van richtlijnen moet gewerkt worden met multi-etnische teams.'

'Testscores spelen een belangrijke rol bij beoordelingen. Zowel in het onderwijs als op de arbeidsmarkt. Het is daarom van het grootste belang dat die tests deugen.'

'Elk kind is uniek. En testscores zijn niet heilig.'

'In Amerika zijn psychologen gerechtelijk te vervolgen, hier niet, dus zijn ze daar verder dan hier, want anders krijgen ze gewoon een proces aan de broek.'

'In de psychiatrie heeft de allochtone deskundigheid meer invloed dan in de psychologie. Er wordt daar gebruik gemaakt van de imam, hindoe priester, winti, etc.'

Haydar Onat, poli psychiatrie, Mesos Medisch Centrum Utrecht

Geldigheid conclusies commissie Hofstee

'Ja, die gelden nog steeds. Er zijn geen revisies van tests geweest.'

'Het is niet alleen etnocentrisme, het is ook hoe de taal is samengesteld in een zin. Dit speelt ook al binnen het Nederlandse taalgebied, tussen verschillende regio's. Met allochtonen zijn de foutmarges in begrip alleen nog maar groter. Men denkt bij de taalbeheersing, ze spreken Nederlands dus ze begrijpen het, maar dat is niet zo.'

'Als je ze strikt zou hanteren, kun je geen enkele test gebruiken. Je moet daar een zekere flexibiliteit weten te hanteren en rekening houden met beschrijvende diagnostiek die je als onderbouwing van je tests kunt gebruiken.'

'De validiteit en betrouwbaarheid van de tests is een academische discussie. De betrouwbaarheid van de meeste tests zijn voor de autochtone groepen goed, de tests zijn goed te gebruiken. Door de verandering van de samenleving is het zinvol om van tijd tot tijd de normgroepen opnieuw aan te passen.'

Genoemde richtlijn commissie Hofstee

'De validiteit en de betrouwbaarheid van de tests is in het geding. Er is tot op heden weinig onderzoek naar gedaan. Van iedere willekeurige test kunnen we stellen dat er geen racistische intentie in de test zit. In ieder geval is dat de bedoeling. Het gaat veel te ver om te stellen dat dit zo zou zijn.'

Bekendheid richtlijnen commissie bij testconstructeurs/-gebruikers/getesten

'Er is hier weinig kennis van genomen. Het is even actueel geweest, maar daarna vergeten. Ik vond wel dat het rapport uit een vreemde hoek kwam. Het is gepubliceerd door het LBR en daarmee wordt het een politieke actie. Dat kwam heel raar over in het begin, maar later begreep ik waarom het LBR zoiets samen met het NIP had gedaan.'

Tekortkoming richtlijnen commissie Hofstee

'De voornaamste tekortkoming is dat er na 10 jaar niets bij is gekomen op gebied van tests voor de allochtone groepen in deze samenleving.'

Waarom heeft dit soort onderzoek niet plaatsgevonden?

'Het staat niet hoog op de agenda van de universiteiten. Het wordt niet echt belangrijk gevonden. En als er geen vraag is, dan is er vaak ook geen aanbod. Er komt wel meer belangstelling voor testonderzoek uit de praktijk, met name in de geestelijke gezondheidszorg zien we ontwikkelingen in deze richting. Ook de arbeids- en organisatiepsychologen zijn zich gaan bezighouden met testen voor de allochtone populatie. Blijkbaar is de vraag naar testen voor deze groep toegenomen op de markt.'

Hoe kan dit soort onderzoek bevorderd worden?

'Er moet een werkgroep komen in samenwerking met de universiteiten en onderwijsinstellingen. Verder moet dit soort projecten financieel ondersteund worden. Op het gebied van intelligentietests, persoonlijkheidstests, loopbaanonderzoek, ontbreekt er veel fundamenteel goed onderzoek en zijn goede tests op dit gebied heel schaars. Het NIP en de COTAN kan hier in een voortrekkersrol spelen. In de opleiding moet dit soort onderwerpen een structureel onderdeel gaan vormen van het curriculum.'

'Er zou een database informatiecentrum moeten komen die informatie verzameld samen met de universiteiten. Dat zou betaald moeten worden door het ministerie van WVS, de gezondheidszorginstellingen en de arbeidsorganisaties bijvoorbeeld.'

'Er komt nu wel meer belangstelling voor testonderzoek uit de praktijk, zodra de vraag meer en dwingend wordt, dan zie dat er hier en daar tests op de markt komen.'

Genoemde gereviseerde of nieuwe test

'De LEM, capaciteitentest voor allochtone populatie. Daarnaast zijn er commerciële bureaus zoals Swets en Zeitlinger die tests ontwikkelen, maar die hebben er belang bij dat die onder eigen beheer worden verkocht. Dit is niet de meest gelukkige oplossing.'

Nieuwe richtlijnen

'Er zou een paragraaf over allochtonen moeten zijn, maar er moeten geen aparte richtlijnen komen voor allochtonen. Als je nastreeft om de ethnocentrische bias zo klein mogelijk te houden en je rekening houdt met mogelijke ethnocentrisme, dan kom je al een heel eind, denk ik.'

Terughoudendheid in interpretatie

'Ik overleg met collega's over rapporten die ik schrijf en vraag om constructieve feedback.'

'Je eigen valkuilen in de gaten houden is even belangrijk. Je moet inl. ook je eigen ethnocentrisme niet uit het oog verliezen. Ook allochtonen hebben een culturele, ethnocentrische bril op soms.'

Terughoudendheid in interpretatie bij anderen

'In wat ik kan overzien van mijn vakgebied, zie ik dat er weinig rekening wordt gehouden met de allochtone populatie. Het gebruik van standaard tests en afgaan op de scores is wijdverbreid.'

Inzet van specifieke deskundigen

Specifieke deskundigen op het gebied van testen van allochtonen zijn er nauwelijks in Nederland, wel zijn er veel ervaringsdeskundigen. In het netwerk wordt er wel gebruik van gemaakt van elkaars deskundigheid'.

Wilma Resing, Universiteit Leiden, O&O-psychologie

Bekendheid met richtlijnen commissie Hofstee

'Door mijn werk (o.a. constructie Kinderintelligentietests) ben ik reeds lang op de hoogte van de richtlijnen.'

'Ik ben tevens lid van de COTAN (Commissie Test Aangelegenheden) en ken ook vanuit deze werkgroep van het NIP de richtlijnen van de commissie omdat de COTAN-richtlijnen aangaande beoordelingen van tests zijn aangescherpt op grond van het rapport van de commissie Hofstee.'

Genoemde richtlijnen commissie Hofstee

'Tests mogen niet zomaar worden afgenomen en gebruik van normen en interpretatie van scores mag zeker niet automatisch gebeuren. Elke test moet gescreend worden op kwetsende opgaven en itembias.'

Bekendheid richtlijnen commissie Hofstee bij testconstructeurs/-gebruikers/getesten

'De testconstructeurs kennen de richtlijnen wel, voor zover het mijn eigen vakgebied – de constructie van intelligentietests – betreft, buiten dit vakgebied kan ik dat niet goed overzien. De constructeurs zouden wel op de hoogte moeten zijn van de richtlijnen, immers de COTAN-beoordelingscriteria zijn op dit punt aangescherpt.'

'Of de gebruikers van de tests de richtlijnen van de commissie allen kennen betwijfel ik, in elk geval binnen de ontwikkelingspsychologie.'

Praktijkdoorwerking richtlijnen commissie Hofstee

'Deze richtlijnen spelen expliciet een rol in het construeren en beoordelen van nieuw ontwikkelde tests. En ik gebruik ze in het onderwijs dat ik geef aan mijn studenten.'

Genoemd (recent) onderzoek

'Naast het onderzoek dat ik zelf met collega's heb uitgevoerd naar de scores van allochtone kinderen op de RAKIT (kinderintelligentietest) en de SCHOB-L-R (beoordelingsinstrument voor schoolgedrag van kinderen), ken ik o.a. het volgende onderzoek. Op A&O-gebied is er het onderzoek van Te Nijenhuis naar de voorspellende waarde van testcores voor allochtone en autochtone werknemers. Daarnaast het onderzoek van Van de Vijver en zijn promovendi naar het ontwikkelen van een heel ander soort test, nl. het meten van o.a. de reactietijd in plaats van het goede antwoord.'

'De onderzoeken die gedaan zijn, zijn tamelijk theoretisch en de resultaten zijn niet altijd direct toepasbaar in de praktijk. Het onderzoek van Van Leest naar persoonlijkheidstests vormt daarop een uitzondering. Hij heeft aparte uitspraken gedaan over allochtone groepen. Dat geldt ook voor het onderzoek van Van de Berg (NOA). De rest is vooral theoretisch van belang en moet zijn praktische uitwerking (voor het testen van allochtonen) nog krijgen.'

Voorstellen voor verder onderzoek

'Er zou longitudinaal onderzoek gedaan moeten worden b.v., naar de acculturatie van groepen, of naar de verandering van testcores over tijd, gecorrigeerd voor verblijfsduur.'

'Ook zouden bijv. de testcores van Turkse kinderen in Turkije en in Nederland in de eigen taal en in de Nederlandse taal tegen elkaar kunnen worden afgezet en op basis daarvan zouden voorschriften voor interpretatie gemaakt kunnen worden, zodat je beter kunt vergelijken. Ook zou het onderzoek naar leerpotentieel verder moeten worden uitgebouwd voor allochtone groepen (omdat daarin niet van de behaalde score maar van de score na training en van de mate waarin van training wordt geprofiteerd wordt uitgegaan).'

'Checken of uitspraken in de handleiding over interpretatie van scores wel kloppen. Of de Turkse taal overheerst of de Nederlandse, hoe lang zijn de onderzochte personen al hier, de mate van acculturatie, dat zou je moeten uitsplitsen. Groeien groepen allochtonen en autochtonen qua testcores naar elkaar toe als je de sociaal-economische factoren constant houdt of en de tijd die men al in Nederland is als variabele meeneemt? Dat is wel de verwachting.'

Waarom heeft dit soort onderzoek niet plaatsgevonden?

'De onderzoekscultuur van de universiteit bemoeilijkt dit soort onderzoek, door de publiciteitsdruk waaraan wetenschappers bloot staan. Over toegepast onderzoek, en zeker testconstructie, valt moeilijk in Engelstalige tijdschriften te publiceren. Een testhandleiding telt niet mee als publicatie.'

'De financiering is ook een probleem. Voor NWO (2^e geldstroom) is dit soort onderzoek te toegepast. En verder is het ook relatief duur om voor Nederland alleen een onderzoek naar een bepaalde test te doen en tot een revisie te komen.'

Hoe kan dit soort onderzoek bevorderd worden?

‘Wat je zou willen is een instelling/ stichting waarin testverkopers, testgebruikers, universiteit en ministerie samen geld poolen om dit soort onderzoek te bekostigen, en een commissie die beoordeeld welk onderzoek leidt tot een verbetering van de situatie. Een onafhankelijke opdrachtgever is daarbij ook cruciaal.’

Criteria voor testkeuze

‘Tijdsdruk en moeilijkheidsgraad spelen ook een rol bij de keuze voor tests. De LIR (test-training-test) waarmee je het leerpotentieel van kinderen test, die ik ontwikkeld heb, duurt 2 tot 3 keer zo lang als een standaard intelligentietest en daar is vaak geen financiering voor. En het is ook iets complexer om de test af te nemen. Datzelfde geldt voor de LEM. Deze tests worden daarom niet zoveel toegepast.’

Nieuwe richtlijnen

‘In voortgaande discussies in bijv. de Psycholoog en andere artikelen aangaande het onderwerp, worden ook richtlijnen en tips gegeven die bij elkaar gesprokkeld kunnen worden.’
‘De onderzoeksbevindingen uit recent onderzoek zijn, ondanks het feit dat ze meestal afkomstig zijn uit het gebied van werving, selectie en loopbaanbegeleiding, wel algemeen toepasbaar.’
‘En verder zijn nog veel vragen en onduidelijkheden over tests en scores, ook bij de testconstructeurs. Er zal eerst meer onderzoek gedaan moeten worden voordat er verdere zinvolle richtlijnen kunnen worden ontwikkeld.’

Wenselijkheid van en eisen aan nieuwe richtlijnen

‘Er moet een vergelijking met de scores van de eigen groep en met de scores van de normeringsgroep mogelijk zijn. Een verdere uitwerking van de paragraaf over ethiek, en het afdwingbaar kunnen verbieden van het afnemen van een test wanneer er geen voorschriften voor interpretatie zijn voor een bepaalde groep. Specifiekere richtlijnen kunnen pas na verder onderzoek geformuleerd worden, want die zijn afhankelijk van de tests, de instrumenten en je domein.’

Terughoudendheid in interpretatie bij anderen

‘De cursussen Interculturele Diagnostiek van het NIP Dienstencentrum worden druk bezocht. Er is dus wel degelijk een bewustzijn van de problematiek en men probeert zich daarover te informeren.’

Inzet van specifieke deskundigen

‘Iedereen met een diagnostiek aantekening moet ook diagnostisch onderzoek bij allochtonen kunnen doen. Maar collega's uit de allochtone groepen zouden wel betrokken kunnen worden bij het interpretatieproces. Als extra paar ogen, of ook om de hele beoordeling samen te doen.’

Raymond Verboom, Altrecht, Geestelijke gezondheidszorg, Divisie Stad Utrecht, afdeling Volwassenen

Geldigheid conclusies commissie Hofstee

‘In grote lijnen gelden ze nog steeds. Uit een pilotstudie van de VKP⁴ bleek dat allochtonen door verkeerdt begrijpen van de items uiteindelijk als zieker gescoord worden dan ze in de praktijk blijken te zijn.’

Criteria voor testkeuze

‘Het opleidingsniveau van de cliënt, de mate van acculturatie en taalbeheersing bepalen welke test ik kies. Meer de praktische en individuele omstandigheden dus dan de specifieke richtlijnen.’

Tekortkoming richtlijnen commissie Hofstee

‘Naast tests zou je ook observatie en interactionele gegevens moeten opnemen in het beoordelingssysteem.’

4 Vereniging voor Klinische Psychotherapie.

Uitgevoerd onderzoek

'Er is voor zover mij bekend geen onderzoek gedaan op mijn eigen werkterrein. Wel op andere gebieden zoals b.v. het onderzoek van Van den Berg van de NOA, die de MCT-M ontwikkeld heeft voor het werkterrein van de werving en selectie.'

Waarom heeft dit soort onderzoek niet plaatsgevonden?

'Met dit soort onderzoek kun je in de universitaire wereld niet scoren. Er is ook geen centrale aansluiting om b.v. het testgebruik in beeld te brengen via een enquête, om eventueel nieuwe tests te construeren. Veel blijft beperkt tot particuliere initiatieven op dit terrein, bijvoorbeeld van ons Platform Psychodiagnostiek bij Migranten en Vluchtelingen, dat nu een lijn heeft met GGZ-Nederland.'

'Er is geen traditie, geen consistente onderzoekslijn vanuit de universiteit in samenwerking met het klinische veld. Er moet meer een brug geslagen worden tussen de universiteit en de praktijk. Als klinisch psycholoog is dit voor mij een ideaal. In de geestelijke gezondheidszorg slokken de curatie en face-to-face gesprekken met de cliënten veel tijd op. Er is daardoor bijna geen tijd voor onderzoek. In de gesubsidieerde praktijk moeten productiecijfers gehaald worden. Managers hebben hier andere belangen dan de mensen op de werkvloer. Dat heb je in de vrije praktijk minder, daar kan meer maatwerk geleverd worden. Maar op de Riaggs wordt er geen consistente kwaliteit geleverd wat dit aangaat. Tenzij de dienstverlening aan allochtonen en research een speerpunt is van het beleid van de vestiging. Dan wordt er (hopelijk) tijd en geld voor vrijgemaakt.'

Voorstellen voor verder onderzoek

'De normen bij vragenlijsten en tests zijn veelal gebaseerd op autochtonen en er is op dit gebied dus gewoon meer gericht onderzoek nodig, zodat je in de praktijk meer verantwoord kunt testen bij allochtonen.'

'Er zou een MTC-L ontwikkeld moeten worden, in navolging van de MCT-M, voor mensen met een lagere opleiding. Juist deze doelgroep vinden we veel in de GGZ. En onderzoek naar het gebruik van projectief materiaal, waarin het narratief van de cliënt naar voren komt. Dat kan helpen bij het helder krijgen van emotionele problemen waar de cliënt zich nog niet expliciet van bewust is.'

Genoemde gereviseerde of nieuwe tests

'Er is in 1996 een Nederlandse Neo PI-R en NEO FFI ontwikkeld op basis van de Big Five. Maar ook bij die "cultuurvrije" tests is bij mijn weten alleen sprake van (jonge) Nederlandse normgroepen. In de praktijk vergroot dat de bruikbaarheid van de tests derhalve nauwelijks. Bovendien kost invulling met behulp van een tolk zeer veel tijd.'

Terughoudendheid in interpretatie

'Overleg met collega's, bijvoorbeeld: wat is gebruikelijk/afwijkend in cultuur X.'

'Daarnaast gebruik je een zo breed mogelijk instrumentarium. B.v. naast tests ook observatie, interactionele gegevens en projectief narratief materiaal. Je probeert door een combinatie van gegevens een zo goed mogelijk beeld te krijgen.'

'Voldoende terughoudendheid wordt bemoeilijkt door het ontbreken van normen toegespitst op allochtonen. Wanneer er alleen Nederlandse normen zijn, die gehanteerd kunnen worden, dan is het twijfelachtig of de terughoudendheid die men probeert te betrachten voldoende is.'

Terughoudendheid in interpretatie bij anderen

'Bij de huidige Gz-psychologen opleiding wordt ernaar gestreefd om grondig diagnostiek te doceren, hoewel er te weinig structurele ruimte is voor interculturele aspecten, behalve in de psychotherapieopleiding van het RINO⁵ in Amsterdam. Maar jonge mensen worden nu in elk geval wel grondig getraind in diagnostiek. Dat werkt ook goed uit voor allochtonen. Als ik daarentegen kijk naar oudere collega's, dan blijken velen niet of onvoldoende opgeleid in testdiagnostiek. En waarmee ik ook niet gelukkig ben, is bijvoorbeeld de volgende wijze van werken: diagnostici doen niet zelf het onderzoek, maar laten dit afnemen door een assistent en baseren dan hun oordeel louter op scores,

⁵ Regionaal Instituut voor Nascholing en Opleiding.

zonder dat zij de cliënt zelf gezien hebben! En vervolgens concluderen ze maar wat. Ook bij autochtonen. Verbijsterend. Het is een verschraving van je vak. Vandaar dat ik niet happig ben op alleen hanteren van de vragenlijstmethode. Er zou een bijscholing moeten plaatsvinden van oudere collega's.'

'En dan heb je nog de relativiteit van je model. De DSM-IV biedt alleen een classificatie, is descriptieve diagnostiek. Als aanvulling heb je een verklaring nodig. De structurele diagnostiek is voor mij veel belangrijker. Borra heeft een aanvulling op de DSM-IV beschreven voor het structureren van je informatie: wat ga je vragen aan je cliënt om cultuur te laten meewegen: Cultural Formulation. Dat zou standaard in de geestelijke gezondheidszorg moeten worden ingevoerd.'

Inzet van specifieke deskundigen

'Je moet geen specifieke deskundigen hiervoor gaan creëren. Het is wel een goede zaak dat medewerkers en collega's ook van multiculturele afkomst zijn. Zo kan er nog eens een extra paar ogen naar kijken vanuit een andere culturele achtergrond. Sowieso moeten alle mensen die diagnostisch werk doen, iets over interculturalisatie in hun opleiding gehad hebben. Tenslotte zijn in het werken met migranten nieuwe psychoanalytische gezichtspunten van belang, die juist de nadruk leggen op een attitude van niet-wetende-nieuwsgierigheid, waarbij er oog is voor het zogenoemde "contextuele zelf".

Romeo Yorks, Yorks & Partners Consultancy

Geldigheid conclusies commissie Hofstee

'Die gelden nog steeds. De situatie is niet veranderd.'

Genoemde richtlijnen commissie Hofstee

'Specifieke deskundigheid nodig, onoordeelkundig gebruik van tests voor allochtonen, bias in wie de test afneemt, interpretatie van testcores en testsamenstelling.'

Bekendheid richtlijnen commissie Hofstee bij testconstructeurs/-gebruikers/getesten

'Testconstructeurs zijn er wel bekend mee. Testgebruikers min of meer. Het zou me niet verbazen als de meesten niet op de hoogte zijn. De kandidaten zeker niet.'

'Testgebruikers gaan denk ik uit van alle verzamelde kennis en informatie, die ze in de loop der jaren hebben opgedaan uit publicaties en artikelen, niet zo zeer van deze richtlijnen expliciet. Het is niet zo dat ze voorbij aan dit punt, nergens rekening mee houden.'

Criteria voor testkeuze

'Ik werk nauwelijks met tests. Tests zijn voor mij niet van allesomvattend belang. Je verliest de persoon uit het oog als je je focust op alleen tests en testconstructie. Dat geldt voor autochtonen zowel als allochtonen. Je moet vooral kijken naar het individu. En is daarbij belangrijker te bekijken hoe iemand functioneert dan hoe iemand is, de procesbenadering. En ook: is het een persoonlijkheidsstoornis of is het delinquente gedrag logisch in het licht van de sociale situatie? Waar wil je voor testen? Dat zijn vragen die ik mij zelf stel. Het zegt mij niets dat iemand een IQ heeft van 98. Wel als iemand een IQ heeft van 50, of ADHD, of als er sprake is van psychopathologie.'

Genoemd uitgevoerd onderzoek

'Er is intelligentieonderzoek gedaan naar aanleiding van dat Amerikaanse boek The Bell Curve. De conclusie was dat er verschil in intelligentie is tussen witten en zwarten.'

Voorstellen voor verder onderzoek

'Onderzoek naar verschillende etnische groepen heeft weinig zin. Er zijn onder de allochtonen erg grote verschillen onderling. Het maakt veel uit of iemand hier al 20 jaar is of nog maar een ½ jaar, zijn ze vernederlandst of traditioneel. Net zo goed als DE Nederlandse cultuur niet bestaat. Ik denk daarom ook dat het onderscheid autochtoon-allochtoon niet zo zinvol is.'

'Mijns inziens zit het dus niet in het mechaniek. Het gaat niet om specifiek onderzoek, maar om de interpretatie en andere dingen daar buiten.'

Genoemde gereviseerde of nieuwe test

'Intelligentietest voor allochtonen.'

Nieuwe richtlijnen

'Ik hecht meer aan een procesbenadering van testen, gericht op beeldvorming van het individu. Richtlijnen helpen mijns inziens niet bij de interpretatie van testresultaten. Het gaat mij om iemands ervaring en vaardigheden in het gebruik van zijn instrumenten in de diagnostiek. En of iemand zich bewust is van zijn eigen inbreng in en invloed op de interactie. Dat zegt mij veel meer dan objectieerbare kennis over het testproces.'

Terughoudendheid in interpretatie

'Iedereen werkt verschillend. Er wordt gestreefd naar uniformiteit. Je werkt samen met andere rapporteurs, van de reclassering, Consultatiebureau voor Alcohol en Drugs – CAD, jeugdpsychiatrie, etc. Ik probeer altijd ik zoveel mogelijk informatiebronnen te gebruiken. Ik wil het dossier zien en de ten-lastelegging. Ik wil informatie over de familie, de sociale omgeving (i.v.m. alternatieve carrière), de persoonlijke historie, eventueel maak ik gebruik van heteroanamnese. Ik vind tests op zich zelf niet zo interessant. Ik heb overigens geen specifieke terughoudendheid ten opzichte van allochtonen. Ik ben überhaupt voorzichtig met mijn oordeel over mensen in het algemeen.'

Terughoudendheid in interpretatie bij anderen

'Ik zie in het algemeen een groot verantwoordelijkheidsgevoel bij (test-)psychologen.'

'In de interpretatie van testcores is ook nog veel onduidelijk, in tegenstrijdige richtlijnen van soms alleen al, b.v. bij de Raven. Iedereen zal ja zeggen op de vraag: houdt u daar rekening mee, maar hoe, dat blijft onduidelijk.'

'Psychologen in de forensische psychologie kunnen in principe werken op hun eigen manier, maar dat de rechter bepaalt welke informatie wel of niet wordt toegelaten in de rapportage. Niet alle rechters laten alle informatie toe over iemands sociale omgeving, persoonlijke historie, familie, acculturatie etc. Er is een uniform rapportagevoorschrift, dat daaraan in principe wel ruimte biedt. De richtlijn is: alle relevante informatie dient gepresenteerd te worden. Maar wat is relevant? Daar heeft de psycholoog dus niet het laatste woord in.

Het wordt in de forensische psychologie overigens wel als een groot probleem ervaren, want in het criminele circuit komt men veel allochtonen tegen.'

In toevoeging op wat besproken werd

'Sommige psychologen gebruiken tests als eeuwig antwoord, want instrumenten zijn het antwoord voor de cultuur waarin wij leven. Een cultuur waarin cognitie belangrijker is dan intuïtie, waar het logisch positivisme de opperste wijsheid is. Ik geloof meer in cognitie EN intuïtie, yin en yang zo je wilt.'

BIJLAGE 2 PERSONALIA VAN DE GEÏNTERVIEWDE DESKUNDIGEN

2

Drs. Paul Abell (1952) is psycholoog en was tot 1994 als senior onderzoeker op het gebied van etnische minderheden aan de vakgroep Arbeids- en Organisationspsychologie van de Universiteit van Amsterdam. Sinds 1994 is hij directeur van EGA HRMConsult, een organisatieadviesbureau dat organisaties adviseert op het bredere terrein van diversiteitsmanagement. Hoofdactiviteiten van EGA zijn beleidsevaluaties, doorlichten van HRM instrumentaria op culturele bias, verrichten van organisatie diagnoses ('diversiteitsaudits') en ontwikkelen van selectie-instrumenten.

Drs. Ernst van Beemen (1953) is sociaal- en organisatiepsycholoog (RU Leiden, 1977). Hij werkt als senior adviseur bij de adviesgroep Bedrijfspsychologie van GITP International BV, vestiging Rotterdam. Als manager en adviseur heeft hij bij verschillende organisaties brede ervaring opgedaan in personeelsmanagement en opleidingen, o.a. m.b.t. internationaal en intercultureel personeelsmanagement (beoordeling, begeleiding en communicatie). Met enkele collega's heeft hij binnen GITP knelpunten en mogelijke oplossingen ('best practices') geïnventariseerd t.a.v. intercultureel assessment en advies.

Drs. Winfried Deijmann (1955) is sinds 1990 internationaal werkzaam als adviseur en gedragstrainer op het gebied van personele mobiliteit, identiteitsontwikkeling en interculturele samenwerking. Het door de Nederlandse socioloog dr. Lex Bos (NPI) ontwikkelde model voor Dynamische Oordeelsvorming en het talentmanagement model van Leo v.d Burg (TRICONA) ligt aan veel van zijn werk ten grondslag. Momenteel is hij verbonden aan de HORIZONGROEP te Bunnik als lid van het MT en als Businessmanager Management Development.

Drs. Hans van Doorn (1953) werkt sinds tien jaar als pedagogisch didactisch medewerker bij het Centrum Educatieve Dienstverlening (CED) in Rotterdam. Hij studeerde psychologie aan de Vrije Universiteit Amsterdam, en werkte daar ook bij de sectie Speciale Pedagogiek.

Dr. Arne Evers (1947) is als universitair hoofddocent werkzaam bij de programmagroep Arbeids- en Organisationspsychologie van de Universiteit van Amsterdam. Geeft onderwijs en doet onderzoek op het gebied van personeelsselectie, discriminatie en vragenlijst- en testontwikkeling. Is tevens lid van de COTAN en mede-auteur van de Documentatie van Tests en Testresearch.

Prof.dr. David Ingleby (1942) volgde zijn opleiding Psychologie aan de University of Cambridge en is daar in 1969 gepromoveerd. Vervolgens heeft hij als onderzoeker bij de Medical Research Council gewerkt en in Cambridge gedoceerd. Vanaf 1984 bekleedde hij de leerstoel Ontwikkeling en Sociale Context aan de Universiteit Utrecht en in 1999 werd hij tot hoogleraar Interculturele Psychologie aan dezelfde universiteit benoemd.

Dr. Wim Kleijn (1948) werkt als universitair docent bij de afdeling psychiatrie van het Leids Universitair Medisch Centrum en als psycholoog onderzoeker bij Centrum '45 in Oegstgeest. Hij heeft onder meer onderzoek gedaan bij allochtone huisartspatiënten, bij vluchtelingen adolescenten in het onderwijs en bij getraumatiseerde vluchtelingen in behandeling bij De Vonk. Daarnaast was hij betrokken bij een enquête onder testassistenten en psychologen over problemen bij het testen van allochtone cliënten.

Mw.drs. Wil Koning (1946) is ontwikkelingspsycholoog/kinder- en jeugdpsycholoog. Heeft vooral ervaring in de onderwijssector, als psychodiagnost, schoolbegeleider en opleider. Momenteel werkzaam bij een pedologisch instituut in Amsterdam (GPI); daar onder meer betrokken bij de ontwikkeling en overdracht van methoden voor handelingsgerichte diagnostiek.

Dr. Jan ter Laak (1944) is universitair hoofddocent aan de capaciteitsgroep Ontwikkelingspsychologie van de Universiteit Utrecht. Hij geeft onderwijs en verricht onderzoek op het gebied van de psychologische diagnostiek van de ontwikkeling, en van de ontwikkeling van de persoonlijkheid. Hij was lid en voorzitter van de Commissie Testaangelegenheden (COTAN), en lid van het Hoofdbestuur van het NIP (als voorzitter van de sector Jeugd).

Drs. Frank van Luijk (1956) studeerde Arbeids- & Organisationspsychologie aan de Vrije Universiteit en Filosofie aan de Universiteit van Amsterdam. Hij is sinds 1983 werkzaam bij LTP, adviseurs voor organisatie- en personeelsvraagstukken, met name op het terrein van assessment en loopbaanadviesing. Sinds 1998 is hij lid van de directie van LTP.

Dr. Bim Mungra (1940) studeerde af in de Psychologie aan de Universiteit van Amsterdam. Hij heeft 26 jaar ervaring met het psychologisch onderzoek bij allochtone en autochtone leerlingen in Amsterdam bij het Advies en Begeleidings Centrum. In 1990 is hij aan de Universiteit van Leiden gepromoveerd op het thema Hindoestaanse gezinnen in Nederland: culturele en sociaal-structurele aspecten van integratie.

Drs. Haydar Onat (1963, Turkije) werkt als onderzoekspsycholoog in een Tbs-Kliniek in Amsterdam en als klinisch psycholoog i.o. op de poli psychiatrie van Mesos Medisch Centrum in Utrecht. Is per 1 augustus 2001 verbonden aan PI Utrecht Wolvenplein als inrichtingspsycholoog. Aandachtsgebieden zijn: (transculturele) diagnostiek, onderzoek en advisering, kortdurende psychologische behandelvormen, interculturalisatie, human resource management & consultancy.

Mw.dr. Wilma C.M. Resing (1952) is als universitair docent verbonden aan de Sectie Ontwikkelings- en Onderwijspsychologie van de Universiteit Leiden. Zij promoveerde in 1990 op een proefschrift getiteld: Intelligentie en leerpotentieel. Zij doceert Cognitieve Ontwikkeling en Intelligentie en coördineert het onderwijs in de psychodiagnostiek. Aandachtsgebieden: de ontwikkeling van intelligentie en meer in het bijzonder het inductief redeneren, de constructie van intelligentietests, onderzoek naar leerpotentieel voor jonge autochtone en allochtone kinderen.

Drs. Raymond Verboom (1952) is Klinisch psycholoog/psychotherapeut en psychoanalyticus, sinds 1979 werkzaam in de GGZ en momenteel o.a. verbonden aan Altrecht, Geestelijke gezondheidszorg, Divisie Stad Utrecht, afdeling Volwassenen. Hij houdt zich hier bezig met curatie en (praktijk)opleiding, en coördineert het psychodiagnostisch testonderzoek. Werken met migranten en vluchtelingen vormt een speerpunt van het beleid. Daarnaast is hij actief in de werkgroep "Psychoanalyse en Multiculturele Samenleving", uitgaande van de Ned.Ver.Psychoanalyse, het Ned. Psychoanalytisch Genootschap en de Ned.Ver. v. Psychoanalytische Psychotherapie.

Drs. Romeo Yorks (1950) studeerde klinische Psychologie in Utrecht. Hij was o.a. oprichter en directeur van het Bureau Alcohol & Drugs, en adviseur voor de Surinaamse regering aangaande preventie en behandeling van verslavingen. Sinds 1997 werkt hij als organisatie-adviseur met enkele partners samen in Yorks & Partners Consultancy te Amersfoort, een bureau gespecialiseerd in veranderings- en diversiteitsmanagement. Daarnaast werkt hij als interim-manager en als vast gerechtelijk deskundige bij de arrondissementsrechtbank van Amsterdam.

BIJLAGE 3

3 VRAGENLIJST 'DESKUNDIGEN OVER HET TESTEN VAN ETNISCHE MINDERHEDEN'

1. Vindt u dat de *conclusies van de commissie* (ethnocentrische inhoud, onmogelijkheid tot vergelijkend oordeel te komen en onduidelijkheid over voorkomen van item- en testbias) nog steeds gelden? Waar baseert u uw oordeel op?
2. Bent u bekend met de *richtlijnen en aanbevelingen* van de commissie? Welke schiet(en) u als eerste te binnen? Zijn deze richtlijnen naar uw oordeel bekend bij de door de commissie aangegeven groepen (testconstructeurs, -gebruikers, kandidaten en andere betrokkenen)? Als u bekend bent met de richtlijnen en aanbevelingen, kunt u dan aangeven op welke wijze u daar rekening mee houdt bij de beoordeling van de kwaliteit, geschiktheid en toepasbaarheid van de tests voor de doelgroep? Kunt u ook aangeven hoe naar uw oordeel anderen op het terrein waarop u deskundig bent/werkzaam bent daar rekening mee houden - op basis van welke criteria worden tests gekozen? Kunt u aangeven in hoeverre er tekortkomingen en beperkingen zijn in de richtlijnen en aanbevelingen? Hoe zouden deze tekortkomingen en beperkingen kunnen worden opgeheven? Als u niet bekend bent met de richtlijnen en aanbevelingen, kunt u de achtergrond daarvan aangeven?
3. Is er naar uw mening sinds het verschijnen van het rapport verder *onderzoek* naar toepasbaarheid van psychologische tests verricht? Zo ja, kunt u aangeven welk onderzoek dat is en wat de belangrijkste uitkomsten van dit onderzoek zijn? Spelen naar uw oordeel de uitkomsten van dit onderzoek een rol in de wijze waarop tests worden toegepast? Vindt u dat er voldoende onderzoek op dit terrein is verricht of dient er nog meer te gebeuren? Zo nee, waaraan is dat te wijten en wat dient er naar uw oordeel te gebeuren om dit te bevorderen?
4. Zijn er naar uw oordeel in de afgelopen periode *tests gereviseerd* met het oog op de vermindering van ethnocentrisme? Kunt u voorbeelden noemen? Zijn er *nieuwe tests* ontwikkeld om met het oog op de gesignaleerde tekortkomingen? Zo ja, kunt u aangeven welke dat zijn? Zo nee, waarom is dat naar uw oordeel niet gebeurd? Wordt bij de nieuwe uitgave van tests naar uw oordeel in voldoende mate aangegeven in hoeverre er sprake is van ethnocentrisme? Houdt u bij eventuele oordelsvorming over bruikbaarheid/toepassingsmogelijkheden rekening met bestaand ethnocentrisme in de test? Kunt u een beeld geven van de wijze waarop in uw werkterrein door anderen (collega's, betrokkenen) met dit probleem wordt omgegaan?
5. Vindt u dat er in de afgelopen jaren voldoende goede en bruikbare *richtlijnen* zijn ontwikkeld voor afname van tests bij allochtonen? Zijn deze richtlijnen algemeen toepasbaar of worden deze alleen in bepaalde omstandigheden/binnen bepaalde organisaties/op bepaalde gebieden gebruikt? Kunt u voorbeelden van dergelijke richtlijnen noemen? Als deze richtlijnen bestaan, houdt u daar dan rekening mee in uw oordelsvorming? Als ze naar uw oordeel niet bestaan, vindt u het dan gewenst om deze te ontwikkelen en aan welke eisen zouden deze dan moeten voldoen?
6. Bent u terughoudend bij de *interpretatie* van tests bij allochtone kandidaten? Zo ja, hoe komt deze terughoudendheid tot uiting? Is deze terughoudendheid naar uw oordeel voldoende? Zo nee, hoe wordt dan te werk gegaan? Wat is de reden daarvan? Kunt u een oordeel geven over de wijze waarop binnen het terrein waarop u werkzaam bent deze terughoudendheid vorm krijgt? Acht u dat voldoende of zou dat anders moeten?
7. Wordt bij de interpretatie van testresultaten gebruik gemaakt van *specifieke deskundigen*? Geldt dit naar uw inzicht op het gehele terrein waarop u werkzaam bent?
8. Wilt u nog iets toevoegen? Is er iets niet besproken dat u belangrijk vindt om op te merken?

BIJLAGE 4 CONCLUSIES EN AANBEVELINGEN UIT HET RAPPORT ‘TOEPASBAARHEID VAN PSYCHOLOGISCHE TESTS BIJ ALLOCHTONEN’ VAN DE COMMISSIE-HOFSTEE

Testscreening: algemene conclusies

1. In geen van de twintig gescreende tests heeft de TSC [Testscreeningscommissie, red.] expliciet racistische inhoud aangetroffen. Dat neemt niet weg dat in andere tests, of in verouderde vormen van de gescreende tests, racistische inhoud kan voorkomen. Met name is dat het geval bij de Taaltests voor Kinderen (TvK) en de nu verouderde versie van de DAT.
2. Etnocentriciteit ten gunste van de dominante cultuur is een bijna universeel verschijnsel in de gescreende tests, met name in verbale tests en vragenlijsten.
3. Bijzonder storend zijn items die etnocentrisme onder woorden brengen, zoals de NPV-items ‘Ik vind dat arme landen zichzelf moeten redden’ en ‘Volgens mij komen de beste bruggenbouwers uit ons land’. De bedoeling van dergelijke items is weliswaar, een met etnocentrisme verwante eigenschap te meten. De uitspraken op zich echter, in hun tendentieuze formulering, kunnen de allochtone respondent zeer onaangenaam treffen. Overigens kunnen ook autochtonen zich storen aan de implicatie dat het bevestigend antwoord blijkbaar “goed” kan zijn (op het voorblad van de NPV staat: “Elk antwoord is goed, als het maar uw eigen mening weergeeft”). Tegen het eventuele argument dat deze redenering leidt tot kortwieken van het persoonlijkheidspsychologisch onderzoek, brengt de TSC in dat de omgang met de respondent grenzen stelt aan de testinhoud, en daarmee eventueel aan de constructen waarop de test zich richt.

4. Voor geen van de onderzochte tests, met uitzondering van de RPDMA-serie, is onderzoek gedocumenteerd waarin de gemiddelde scores van allochtonen en autochtonen worden vergeleken. Dit maakt de tests sterk beperkt toepasbaar (zie 2.3, Richtlijnen). Bij de RPDMA-serie werd tussen allochtonen en autochtonen van vergelijkbare vooropleiding een gemiddeld verschil gevonden van ongeveer één standaarddeviatie in het nadeel van de allochtone groep. Ten aanzien van deze batterij is niet aangetoond dat niet van ondervoorspelling sprake is ten opzichte van een unbiased criterium, zodat de conclusie van sterk beperkte toepasbaarheid ook hier geldt. Met nadruk wijst de TSC op de implicatie dat op de scores van de door haar gescreende tests geen vergelijkende oordelen of beslissingen kunnen worden gebaseerd tussen allochtone en autochtone kandidaten voor een functie of opleiding. De TSC is van mening dat vergelijkende research hoge prioriteit verdient. Daarbij valt op te merken dat de situatie in Nederland gezien het recente karakter van de immigratie niet volledig vergelijkbaar is met die in de Verenigde Staten, zodat Amerikaanse researchuitkomsten, met name betreffende de afwezigheid van ondervoorspelling of aanwezigheid van overvoorspelling, niet zonder meer kunnen worden overgebracht op Nederlandse allochtonen.

Wellicht ten overvloede wijst de TSC erop dat ook niet-verbale tests biased kunnen zijn. Tijdens haar screening-activiteiten is de TSC gestuit op de categorie van in dit opzicht ‘quasi-onschuldige’ tests. Een voorbeeld is het nazeggen van reeksen cijfers (vergelijk het reproduceren van een gedictieerd telefoonnummer). Op het oog is dit een niet-verbale test. Een plausibele hypothese is echter dat iedere respondent die deze taak in zijn of haar tweede taal zou moeten uitvoeren, in het nadeel zou zijn vanwege de noodzaak of verleiding, een dubbele vertaalslag uit te voeren: één bij het inprenten en één bij het reproduceren. Een ander voorbeeld is het soort snelheid-en-nauwkeurigheidstests dat bestaat uit het vergelijken van verschillend of gelijk gespelde namen en adressen. Ook dit is geen verbale test, in de zin dat het hier niet om betekenissen gaat maar louter om het vergelijken van symbolen. Niettemin is het plausibel dat de autochtoon vanwege een langer getraind woordbeeld in het voordeel is wanneer de namen en adressen Nederlands zijn.

Onderzoekers hebben veelvuldig geconstateerd dat allochtonen niet alleen op verbale tests lager scoren. De hier gegeven voorbeelden leveren daarvoor min of meer concrete verklarings-

hypothese die in experimenteel onderzoek kunnen worden getoetst, overigens zonder dat de herkomst van de proefpersonen daarbij relevant is. De TSC beveelt zulk onderzoek in de aandacht aan.

5. Bij geen van de gescreende tests is onderzoek naar item bias gedaan. De TSC heeft in de loop van haar screeningswerkzaamheden veelvuldig vermoedens van item bias (en test bias) geuit. Deze vermoedens zijn ook exemplarisch gerapporteerd bij de afzonderlijke tests. Ze dienen uitdrukkelijk als hypothesen te worden opgevat, niet als oordelen die de test zouden diskwalificeren. Uit de empirische literatuur is bekend dat item bias een hachelijk onderwerp is, en dat met name vermoedens omtrent item bias veelvuldig ongegrond blijken.

Aanbevelingen

1. De TSC beveelt de door haar ontwikkelde Richtlijnen (par. 2.3 van dit rapport) en de toepassingen daar van (par. 2.4) in de aandacht van testconstructeurs, -gebruikers, sollicitanten en andere betrokkenen aan.
2. De TSC heeft kennis genomen van de uitspraak van de Commissie Testaangelegenheden van het Nederlands Instituut van Psychologen "... dat het zeer goed mogelijk lijkt de criteria die de test-screeningscommissie heeft gehanteerd in te bouwen in het beoordelingssysteem voor tests bij de op handen zijnde herziening hiervan. Tevens lijkt het mogelijk de criteria op te nemen in een eventuele herdruk van de Richtlijnen voor Psychologische Tests en Studietoetsen". De TSC steunt deze voornemens.
3. De TSC constateert een ernstige achterstand in onderzoek naar test bias in Nederland. Zij beveelt zulk onderzoek in de aandacht van testresearchers en onderzoekfinancierende instanties aan.
4. Vanwege hun ethocentrische inhoud dienen zeer veel Nederlandse tests of subtests grondig te worden gereviseerd.
5. Het ontwikkelen van richtlijnen voor de afname en interpretatie van tests bij sollicitanten ging voorbij de taak van de TSC. In het verlengde van hetgeen zij hierover niettemin heeft opgemerkt in par. 1.2 en 2.3, en van wat anderen hierover hebben geschreven, beveelt de TSC de ontwikkeling van afzonderlijke richtlijnen hiervoor aan.
6. Op basis van de screening van twintig tests beveelt de TSC grote terughoudendheid aan bij toepassing van de voorhanden psychologische tests in het algemeen bij allochtone sollicitanten. Automatische oordeels- of besluitvorming, zonder tussenkomst van een terzake ervaren deskundige, acht de TSC hier ontoelaatbaar.
7. De afname van psychologische tests bij allochtonen vereist afzonderlijke deskundigheid.

BIJLAGE 5 DE TESTS VAN HET RAPPORT 'TOEPASBAARHEID VAN PSYCHOLOGISCHE TESTS BIJ ALLOCHTONEN' VAN DE COMMISSIE-HOFSTEE

Differentiële Aanleg Testserie (DAT)
Test voor Niet Verbale Abstractie (TNVA)
Numerieke Aanleg Testserie (NAT)
Verbale Aanleg Testserie (VAT)
Raven's Progressive Matrices (PM)
Technisch Inzicht (TI)
De MA-testserie van de Rijkspsychologische Dienst (RPD)
General Aptitude Test Battery (GATB)
Wechsler Intelligence Scale for Children Revised (WISC-R)
Wechsler Adult Intelligence Scale (WAIS)
Groninger Intelligentie Test (GIT)
Amsterdamse Beroepen Interesses Vragenlijst (ABIV)
Kuder Beroeps Voorkeuronderzoek
Beroepen Interesse Test (BIT)
Nederlandse Persoonlijkheden Vragenlijst (NPV)
Amsterdamse Biografische Vragenlijst (ABV)
Kinder-ABV (ABV-K en KABV-K)
Prestatie Motivatie Test (PMT)
PMT-K
RPD-Vragenlijst (RPDV)

Sinds het verschijnen van het rapport van de commissie-Hofstee, zijn de items van de GATB (vooral op seksistisch terrein) aangepast en is de BIT gereviseerd. De WISC-R wordt nu gereviseerd, de revisie zal in 2002 verschijnen. Verder is geen enkele test veranderd.

BIJLAGE 6 OVERZICHT VAN IN DE INTERVIEWS GENOEMDE TESTS

6

Afkorting	Referentie
ABV	Amsterdamse Biografische Vragenlijst (1963) van G.J.S. Wilde. Wilde, G.J.S. (1963). <i>Neurotische labiliteit, gemeten volgens de vragenlijstmethode</i> (2 ^e vermeerderde uitgave, 1970). Amsterdam: Van Rossen. Lisse: Swets & Zeitlinger.
DAT ('83)	Differentiële Aanleg Testserie (1983/1991), bewerking van de Differential Aptitude Tests forms S & T (1974) van G.K. Bennett, H.G. Seashore en A.G. Wesman, door A. Evers en W. Lucassen. Evers, A. & Lucassen, W. (1991). <i>Handleiding DAT '83. Differentiële Aanleg Testserie</i> . Lisse: Swets & Zeitlinger.
GATB	General Aptitude Test Battery, GATB versie B 1002-B (1971/1994) van het United States Employment Service (1958), voor Nederlands bewerkt door de Stichting GATB-Research. Flier, H. van der & Boomsma-Suerink, J.L. (1994). <i>Handboek GATB B 1002-B</i> (2 ^e druk). Separate uitgave Normen GATB B 1002-B. Utrecht: Stichting GATB Research.
GEPPS	speciale GTP-versie (verkrijgbaar bij Swets & Zeitlinger) van de Edward Personal Preference Schedule (EPPS) (1973) van A.L. Edward (1967), bewerkt door A.S.H. Tjoa. Tjoa, A.S.H. (1993). <i>EPPS Herzienne Handleiding</i> . Nederlandse bewerking van Edwards Personal Preference Schedule. Lisse: Swets & Zeitlinger.
GIVO	Groninger Intelligentietest voor Voortgezet Onderwijs (1994) van H. van Dijk en P.J. Tellegen. Dijk, H. van & Tellegen, P.J. (1994). <i>Handleiding, Testboekje, Instructieboekje, GIVO, Groninger Intelligentietest voor Voortgezet Onderwijs</i> . Lisse: Swets & Zeitlinger.
GORDON	Gordon Personal Profile and Personal Inventory, niet te verkrijgen op de markt, rechten liggen bij een stichting.
IBO Differentiatie- test	IBO Differentiatietest (1979) van M.J. Koornstra, N.M.E. Neuwahl en W. van Hoorn. Koornstra, M.J., Neuwahl, N.M.E. & Hoorn, W. van. (1979). <i>Handleiding IBO Differentiatietest</i> . Lisse: Swets & Zeitlinger.
LEM	Leertest voor etnische minderheden (1991) van J.H.M. Hamers, M.G.P. Hessels, en J.E.H. van Luit. Hamers, J.H.M., Hessels, M.G.P. & Luit, J.E.H. van. (1991). <i>Handleiding Leertest voor Etnische Minderheden</i> . Lisse: Swets & Zeitlinger.
LIR	Leerpotentieel test voor Inductief Redeneren (1990) van W.C.M. Resing, experimentele versie. De test staat beschreven in Resing, W.C.M. (1990). <i>Intelligentie en leerpotentieel van jonge leerlingen uit het basis- en speciaal onderwijs</i> . Lisse: Swets & Zeitlinger.
MCT-M	Multiculturele Capaciteiten Test – Middelbaar Niveau (1997/1999) van N. Bleichrodt en R.H. van den Berg. Bleichrodt, N. & Berg, R.H. van den. (1999). <i>Handleiding MCT-M, Multiculturele Capaciteiten Test – Middelbaar niveau</i> . Amsterdam: NOA/VU.
MPV	Multiculturele Persoonlijkheds Vragenlijst (1998) van K.I. van der Zee en J.P. van Oudenhoven. Experimentele versie. Zee, K.I. van & Oudenhoven, J.P. van (1999). <i>De Multiculturele Persoonlijkheds Vragenlijst. Vragenlijst voor Multiculturele Effectiviteit. Handleiding</i> . Groningen: RUG, vakgroep Psychologie.
NDT	Nederlandse Differentiatie Testserie (1999), experimentele versie van W. van Hoorn, J. Meijer, R. Oostdam, W. den Brinker, S. van der Sluis en L. van Zoelen. Hoorn, W. van, Meijer, J. & Oostdam, R. (in voorb.). <i>Handleiding bij de Nederlandse Differentiatie Testserie, NDT</i> . Amsterdam: UvA, SCO-Kohnstamm Instituut.

- NEO FFI
Neo PI-R** NEO Persoonlijkheidsvragenlijsten, NEO PI-R en NEO FFI (1996), Nederlandstalige versie van de NEO PI-R Personality Inventory en de NEO FFI Five Factor Inventory van P.T. Costa & R.R. McCrae (1989/1992), door H.A. Hoekstra, J. Ormel en F. de Fruyt.
Hoekstra, H.A., Ormel, J. & De Fruyt, F. (1996). *NEO PI-R – NEO FFI. Big Five Persoonlijkheidsvragenlijsten. Handleiding*. Lisse: Swets & Zeitlinger.
- NPV** Nederlandse Persoonlijkheids Vragenlijst (1974/1985) van F. Luteijn, J. Starren en H. van Dijk.
Luteijn, F., Starren, J. & Dijk, H. van. (1985). *Handleiding Nederlandse Persoonlijkheids Vragenlijst* (herziene uitgave). Lisse: Swets & Zeitlinger.
- RAKIT** Revisie Amsterdamse Kinder Intelligentie Test (1984/1987) van N. Bleichrodt, P.J.D. Drenth, J.N. Zaal en W.C.M. Resing.
Bleichrodt, N., Drenth, P.J.D., Zaal, J.N. & Resing, W.C.M. (1984). *Revisie Amsterdamse Kinder Intelligentie Test. Instructie, normen, psychometrische gegevens*. Lisse: Swets & Zeitlinger.
- Raven – PM** Progressive Matrices, PM (1938) van J.C. Raven.
Raven, J.C., Court J.H. & Raven, J. (1979). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. Section I General Overview. London: H.K. Lewis & Co.
- Raven – CPM** Coloured Progressive Matrices, CPM (1947) van J.C. Raven, bewerking en normering door W.H.J. van Bon, 1986.
Bon, W.H.J. van. (1986). *Raven's Coloured Progressive Matrices. Nederlandse normen en enige andere uitkomsten van onderzoek*. Lisse: Swets & Zeitlinger.
- SCHOBL-R** Schoolgedrag Beoordelingslijst, SCHOBL-R (1980/1993), revisie van de SCHOBL (1980) van J. Zaal door N. Bleichrodt, W.C.M. Resing en J.N. Zaal.
Bleichrodt, N., Resing, W.C.M. & Zaal, J.N. (1993). *Beoordeling van Schoolgedrag, SCHOBL-R Handleiding en Verantwoording*. Lisse: Swets & Zeitlinger.
- SCL-90** Klachtenlijst, SCL-90 (1981/1986), bewerking van de Symptom Checklist van L.R. Derogatis, R.S. Lipman & L. Covi (1973) door W.A. Arrindell en J.H.M. Ettema.
Arrindell, W.A. & Ettema, J.H.M. (1986). *Handleiding bij een multidimensionale psychopathologie-indicator*. Lisse: Swets & Zeitlinger.
- SON-R 2½-7** Snijders-Oomen Niet-Verbale intelligentietest – SON-R 2½-7 (1998), van P.J. Tellegen, M. Winkel, B.J. Wijnberg-Williams en J.A. Laros.
Tellegen, P.J., Winkel, M., Wijnberg-Williams, B.J. & Laros, J.A. (1998). *Snijders-Oomen Niet-verbale Intelligentietest SON-R 2½-7. Handleiding en Verantwoording*. Lisse: Swets & Zeitlinger.
- SON-R 5½-17** Snijders-Oomen Niet-Verbale intelligentietest – SON-R 5½-17 (1989) van J.Th. Snijders, N. Snijders-Oomen, J.A. Laros, M.A.H. Huijnen en P.J. Tellegen.
Snijders, J.Th., Tellegen, P.J. & Laros, J.A. (1988). *Snijders-Oomen niet-verbale intelligentietest SON-R 5½ -17. Verantwoording en Handleiding*. Groningen: Wolters-Noordhoff.
- TAK** Taaltoets Allochtone Kinderen (1986), van L. Verhoeven en A. Vermeer.
Verhoeven, L. & Vermeer, A. (1986). *Taaltoets Allochtone Kinderen, Lessuggesties*. Tilburg: Zwijzen.
- TvK** Taaltests voor Kinderen (1982) van W.H.J. van Bon.
Bon, W.H.J. van. (1982). *Handleiding Taaltests voor Kinderen, TvK*. Lisse: Swets & Zeitlinger.
- WISC-R** Wechsler Intelligence Scale for Children – Revised (1986), bewerking van de WISC-R 1974 van D. Wechsler door de WISC-R Projectgroep.
WISC-R Projectgroep: P.P. van Haasen, E.E.J. De Bruyn, Y.J. Pijl, Y.H. Poortinga, H.C. Spelberg, G. Van der Steene, P. Coetsier, R. Spoelders-Claes & J. Stinissen (1986). *WISC-R, Wechsler Intelligence Scale for Children – Revised, Nederlandstalige uitgave. Deel I. Testinstructie; Deel II. Scoring en Normen; Deel III. Verantwoording*. Lisse: Swets & Zeitlinger.

BIJLAGE 7 LITERATUUR

- American Educational Research Association (AERA), American Psychological Association (APA) en National Council on Measurement in Education (NCME). (1999). *The Standards for Educational and Psychological Testing*. Washington, DC: EARA.
- Berg, R.H. van den. (2001). *Psychodiagnostiek bij kandidaten met verschillende culturele achtergronden*. Amsterdam: Vrije Universiteit, Amsterdam: NOA.
- Bleichrodt, N. & van de Vijver, F.J.R. (Red.) (2001). *Diagnostiek bij allochtonen. Mogelijkheden en beperkingen van psychologische tests*. Lisse: Swets & Zeitlinger
- Burg, L. van der. (1999). *Doen waar je goed in bent. Menselijke kwaliteit in organisaties*. Amsterdam: Addison Wesley Longman.
- Borra, R. (1998). Cultuur en de DSM-IV. *Maandblad Geestelijke Volksgezondheid*, 53 nr. 11, 1086-1102.
- Diagnostic and Statistical Manual of Mental Disorders Fourth Edition*. (1994). Washington, DC: American Psychiatric Association.
- Evers, A., Vliet-Mulder, J.C. van & Groot, C.J. (2000). *Documentatie van tests en testresearch in Nederland*. Deel I Testbeschrijvingen; Deel II Testresearch. Assen: Van Gorcum & Comp., Amsterdam: NIP Dienstencentrum.
- Helms-Lorenz, M. (2001). *Assessing cultural influences on cognitive test performance: A study with migrant children in the Netherlands*. Tilburg : Tilburg University
- Herrnstein, R.J. & Murray, C. (1994). *The Bell Curve: Intelligence and class structure in American life*. New York: Free Press.
- Hessels, M.G.P. (1993). *Leertest voor Etnische Minderheden. Theoretische en empirische verantwoording*. Rotterdam: RISBO.
- Hofstee, W.K.B., Campbell, W.H., Eppink, A., Evers, A, Joe, R.C., Koppel, J.M.H. van de, Zweers, H., Choenni, C.E.S. & Zwan, T.J. van der. (1990). *Toepasbaarheid van psychologische tests bij allochtonen*. LBR-reeks, nr. 11. Rotterdam: Landelijk Bureau Racismebestrijding.
- Hofstee, W., Berge, J. ten, & Drenth, P. (1998). Intelligentie en herkomst. Een normatieve analyse. *De Psycholoog*, 10, 451-454.
- Internationale Richtlijnen voor het Gebruik van Tests* (in druk⁶), Belgische Federatie van Psychologen, Nederlands Instituut van Psychologen, Internationale Test Commissie (ITC).
- Leest, P.F. van. (1997). *Persoonlijkheidsmeting bij allochtonen*. Amsterdam: Vrije Universiteit.
- Nijenhuis, J. te (1997). *Comparability of test scores for immigrants and majority group members in the Netherlands*. Amsterdam: Vrije Universiteit.
- Resing, W.C.M. & Hessels, M.G.P. (2001). Het meten van de cognitieve mogelijkheden en het schoolgedrag van allochtone kinderen. In N. Bleichrodt & F. van de Vijver (Red.), *Diagnostiek bij allochtonen. Mogelijkheden en beperkingen van psychologische tests*. Lisse: Swets & Zeitlinger.

Literatuur uit *De Psycholoog*

- Berg, R. van den & Leest, P. van. (1999). Praktisch testgebruik bij allochtonen. Wanneer zinnig, nuttig of waardeloos? *De Psycholoog*, 6, 256-260.
- Evers, A. & Nijenhuis, J. te. (1999). Liever speciale dan traditionele cognitieve capaciteitentests voor allochtonen? Een vergelijking. *De Psycholoog*, 6, 250-255.
- Tellegen, P. (2000). Verantwoord testgebruik bij allochtonen. Een reactie. *De Psycholoog*, 5, 231-235.

⁶ Voor meer informatie: www.intestcom.org.

BIJLAGE 8 SUGGESTIES, PER WERKVELD EN ALGEMEEN



Werving en Selectie

Onderzoek naar/bij:

- criteriumbias;
- de relatie tussen test- en criteriumscores voor de verschillende groepen;
- voorspellende waarde;
- de werving en selectie van de politie zou onderzoek naar validiteit en voorspellende waarde van tests moeten doen. Zij hebben veel ervaring met het testen van allochtonen.

Tips voor testgebruik:

- aan capaciteitentests een zwaarder gewicht toekennen in selectieprocedures vanwege hun hogere validiteit;
- assessment veel breder maken: niet alleen tests maar een 360 graden beoordeling, o.a. met rollenspellen, interviews. Er moeten daarbij ook nieuwe instrumenten ontwikkeld worden, rollenspellen zijn ook ethocentrisch;
- maak veel gebruik van goed voorbereide en gestructureerde interviews, gericht op competenties (eerdere ervaringen) en cultuurverschillen;
- benut alle indrukken uit assessment-simulaties (rollenspellen) ook als aanvulling op andere assessment onderdelen;
- bij een 'Fact Finding' krijgt men al snel een indruk van hoe probleemanalyse en oordeelsvorming verloopt. Deze indicatie voor informatieverwerking kan gelegd worden naast de indrukken uit klassieke intelligentietests;
- maak vooral gebruik van herkenbare en voor de vraagstelling relevante situaties en simulaties. Maak geen gebruik van postbakken bij mensen die de Nederlandse taal te weinig beheersen. Lezen, begrijpen en beantwoorden kosten tijd en energie en leveren ruis op.

Geestelijke gezondheidszorg

Onderzoek naar/bij:

- onderzoek naar de SCL-90;
- ontwikkeling van een MCT-L voor mensen met lagere opleidingsniveaus.

Tips voor testgebruik:

- observatie in het beoordelingssysteem betrekken;
- gebruik algemene brede-band-persoonlijkheidsvragenlijsten (bijv. GEPPS) en geen specifiek gerichte (bijv. NPV);
- Borra heeft een aanvulling op de DSM-IV beschreven voor het structureren van je informatie: wat ga je vragen aan je cliënt om cultuur te laten meewegen: Cultural Formulation. Dat zou standaard in de GGZ moeten worden ingevoerd.

Onderwijs

Onderzoek naar/bij:

- bij de gebruikers zijn de richtlijnen en aanbevelingen van de commissie te weinig bekend. Het CITO zou dat kunnen doen, maar dat is niet gebeurd;
- lange termijnonderzoek naar juistheid van doorverwijzingen in het onderwijs;
- onderzoek naar de CITO-woordenschatstest voor allochtone kinderen;
- opnieuw normeren van de RAKIT en WISC, door het CITO;
- de testscores van Turkse kinderen in Turkije en in Nederland in de eigen taal en in de Nederlandse taal zouden tegen elkaar kunnen worden afgezet, en op basis daarvan zouden voorschriften voor interpretatie gemaakt kunnen worden, zodat je beter kunt vergelijken;
- het onderzoek naar leerpotentieel verder uitgebouwen voor allochtone groepen (omdat daarin niet van de behaalde score maar van de score na training en van de mate waarin van training wordt geprofiteerd wordt uitgegaan).

Tips voor testgebruik:

- multi-etnische onderzoeksteams bij schoolbegeleidingsdiensten.

ALGEMENE SUGGESTIES

Onderzoek naar/bij:

- bij vluchtelingen;
- naar de grenzen van de invloed van verblijfsduur;
- longitudinaal onderzoek, b.v. naar de acculturatie van groepen, of naar de verandering van testcores over tijd, gecorrigeerd voor verblijfsduur;
- naar observatie;
- checken of uitspraken in de handleiding over interpretatie van scores wel kloppen. Of de Turkse taal overheerst of de Nederlandse, hoe lang zijn de onderzochte personen al hier, de mate van acculturatie, dat zou je moeten uitsplitsen. Groeien groepen allochtonen en autochtonen qua testcores naar elkaar toe als je de sociaal-economische factoren constant houdt of en de tijd die men al in Nederland is als variabele meeneemt? Dat is wel de verwachting;
- naar persoonlijkheidstest, de Big Five naar culturele achtergrond uitsplitsen voor adolescenten b.v.;
- onderzoek á la Leo van der Burg: Doen waar je goed in bent, iedereen deugt, maar waarvoor;
- naar intelligentietests;
- naar accuratesse van uitspraken in de handleiding over interpretatie van resultaten/scores;
- opnieuw de 20 meest gebruikte tests screenen;
- verdere richtlijnen, b.v. aan de hand van de *Standards '99*;
- het bijhouden van instrumenten.

Onderzoek door:

- allochtone testconstructeurs;
- multi-etnische teams van testconstructeurs;
- multi-etnische onderzoeksteams voor richtlijnontwikkeling.

Tips voor testgebruik:

- allochtone beoordelaars en psychologen ook laten kijken naar de resultaten;
- gebruik tests met een duidelijk, eenduidig en eenvoudig taalgebruik;
- maak een aparte rubriek in de testbeoordeling: in hoeverre is sprake van ethnocentrisme in de test, b.v. aan de hand van de *Standards '99*;
- Nederlandse spreekvaardigheidstest bij vluchtelingen standaard toepassen. Voorschrijven om geen tests af te nemen als het Nederlands te beperkt is;
- gebruik bij het meten van intelligentie een breed scala van soorten testen, die zowel 'fluid' als 'crystallized' intelligentie meten, en focus daarbij op de 'G-factor';
- gebruik naast 'speed' tests ook of liever 'power' tests;
- maak gebruik van de Multiculturele Capaciteiten Test (MCT-M of -H) om de hierbij behorende aanvullende hulpmiddelen en referentiegroepen te kunnen gebruiken;

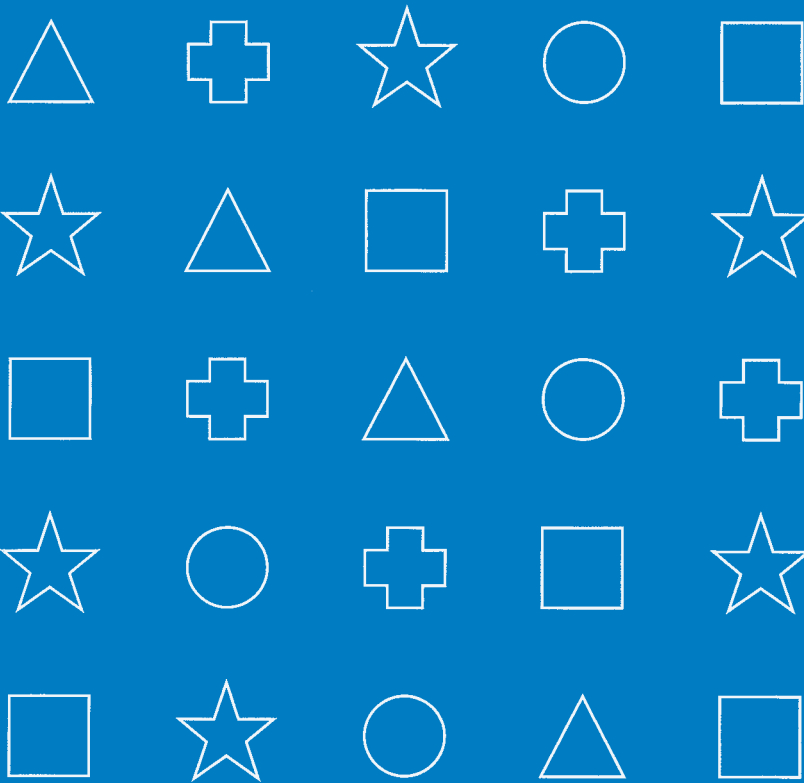
- geef uitgebreide instructie en ken enkele woorden uit hun eigen taal om kandidaten op hun gemak te stellen;
- in voortgaande discussies in bijv. de Psycholoog en andere artikelen aangaande het onderwerp, worden ook richtlijnen en tips gegeven die bij elkaar gesprokkeld kunnen worden;
- er moet een vergelijking met de scores van de eigen groep en met de scores van de normeringsgroep mogelijk zijn;
- een verdere uitwerking van de paragraaf over ethiek, en het afdwingbaar kunnen verbieden van het afnemen van een test wanneer er geen voorschriften voor interpretatie zijn voor een bepaalde groep;
- uitbrengen van een bundel die regelmatig wordt ge-update met de ervaringen uit de praktijk, zoals de huisartsen protocollen kennen, met informatie over welke tests het beste werken in welke situaties. Dat geeft sturing aan het gebruik van de tests. Op welke gronden er nu gekozen wordt voor welke test is onduidelijk.

Scholing van testgebruikers:

- Nederlandse psychologen moeten geschoold worden in interculturele setting;
- aparte scholing voor het testen van allochtonen door een klein bureau;
- bredere scholing assessoren;
- de oudere collega's dienen bijgeschoold te worden.

Tips voor een onafhankelijke instelling

- Een instelling/stichting waarin testverkopers, testgebruikers, universiteit en ministerie samen geld poolen om dit soort onderzoek te bekostigen, en een commissie die beoordeelt welk onderzoek leidt tot een verbetering van de situatie. Een onafhankelijke opdrachtgever is daarbij ook cruciaal;
- Formering van een samenwerkingsverband tussen universiteit en de praktische mensen samen, met geld van de overheid om onderzoek op dit vlak te laten uitvoeren;
- Formeren van een werkgroep in samenwerking met de universiteit voor het doen van onderzoek. NIP of COTAN zouden een voortrekkersrol moeten spelen bij het op de agenda krijgen van onderzoek naar tests bij allochtonen in de geestelijke gezondheidszorg;
- Formering van een werkgroep of platform ter coördinatie en informatie om aanbevelingen op dit vlak te kunnen doen. Er is behoefte aan een centraal punt om ontwikkeling van richtlijnen te sturen, misschien door het NIP, als dat maar niet betekent dat alleen de leden bereikt worden. Er zou een database informatiecentrum moeten komen dat informatie verzamelt samen met de universiteiten. Dat zou betaald moeten worden door het ministerie van OC&W, gezondheidszorg en arbeidsorganisaties.



Für bestimmte Zwecke
als erforderlich erw
Tabelle 2.9 beschreib
Die Formulierungen
des amerikanischen